PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/29, C07K 13/00, 15/10
C12N 1/21, C07K 7/04

A2

(11) International Publication Number: WO 94/11512

(43) International Publication Date: 26 May 1994 (26.05.94)

(21) International Application Number: PCT/US93/11000

(22) International Filing Date: 12 November 1993 (12.11.93)

(30) Priority data:
07/975,179 12 November 1992 (12.11.92) US

(71) Applicant: IMMULOGIC PHARMACEUTICAL CORPORATION [US/US]; 610 Lincoln Street, Waltham, MA 02154 (US).

(72) Inventors: KUO, Mei-chang ; 5 Cox Road, Winchester, MA 01890 (US). YEUNG, Siu-Mei, Helena ; 2 Kimball Court, Apartment 303, Woburn, MA 01801 (US). BRAUER, Andrew, W. ; 21 Gedney Court, Salem, MA 01970 (US). POLLOCK, Joanne ; 51 Newcomb Street, Arlington, MA 02174 (US).

(74) Agents: YANKWICH, Leon, R.; ImmuLogic Pharmaceutical Corporation, 610 Lincoln Street, Waltham, MA 02154 (US) et al.

(81) Designated States: AU, CA, JP, KR, NZ, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: ALLERGENIC PROTEINS AND PEPTIDES FROM JAPANESE CEDAR POLLEN

(57) Abstract

The present invention provides nucleic acid sequences coding for the *Cryptomeria japonica* major pollen allergen *Cry j* II and fragments thereof. The present invention also provides purified *Cry j* II and at least one fragment thereof produced in a host cell transformed with a nucleic acid sequence coding for *Cry j* II or at least one fragment thereof and fragments of *Cry j* prepared synthetically. *Cry j* II and fragments thereof are useful for diagnosing, treating, and preventing Japanese cedar pollinosis.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

ALLERGENIC PROTEINS AND PEPTIDES FROM JAPANESE CEDAR POLLEN

Background of the Invention

Genetically predisposed individuals, who make up about 10% of the population, become hypersensitized (allergic) to antigens from a variety of environmental sources to which they are exposed. Those antigens that can induce immediate and/or delayed types of hypersensitivity are known as allergens. (King, T.P., *Adv. Immunol.* 23: 77-105, (1976)). Anaphylaxis or atopy, which includes the symptoms of hay fever, asthma, and hives, is one form of immediate allergy. It can be caused by a variety of atopic allergens, such as products of grasses, trees, weeds, animal dander, insects, food, drugs, and chemicals.

The antibodies involved in atopic allergy belong primarily to the IgE class of immunoglobulins. IgE binds to mast cells and basophils. Upon combination of a specific allergen with IgE bound to mast cells or basophils, the IgE may be cross-linked on the cell surface, resulting in the physiological effects of IgE-antigen interaction. These physiological effects include the release of, among other substances, histamine, serotonin, heparin, a chemotactic factor for eosinophilic leukocytes and/or the leukotrienes, C4, D4, and E4, which cause prolonged constriction of bronchial smooth muscle cells (Hood, L.E. et al. *Immunology* (2nd ed.), The Benjamin/Cumming Publishing Co., Inc. (1984)). These released substances are the mediators which result in allergic symptoms caused by a combination of IgE with a specific allergen. Through them, the effects of an allergen are manifested. Such effects may be systemic or local in nature, depending on the route by which the antigen entered the body and the pattern of deposition of IgE on mast cells or basophils. Local manifestations generally occur on epithelial surfaces at the location at which the allergen entered the body. Systemic effects can include anaphylaxis (anaphylactic shock), which is the result of an IgE-basophil response to circulating (intravascular) antigen.

Japanese cedar (Sugi; *Cryptomeria japonica*) pollinosis is one of the most important allergic diseases in Japan. The number of patients suffering from this disease is on the increase and in some areas, more than 10% of the population are affected. Treatment of Japanese cedar pollinosis by administration of Japanese cedar pollen extract to effect hyposensitization to the allergen has been attempted. Hyposensitization using Japanese cedar pollen extract, however, has drawbacks in

that it can elicit anaphylaxis if high doses are used, whereas when low doses are used to avoid anaphylaxis, treatment must be continued for several years to build up a tolerance for the extract.

The major allergen from Japanese cedar pollen has been purified and designated as Sugi basic protein (SBP) or *Cry j* I. This protein is reported to be a basic protein with a molecular weight of 41-50 kDa and a pI of 8.8. There appear to be multiple isoforms of the allergen, apparently due in part to differential glycosylation (Yasueda et al. (1983) *J. Allergy Clin. Immunol.* 71: 77-86; and Taniai et al. (1988) *FEBS Letters* 239: 329-332. The sequence of the first twenty amino acids at the N-terminal end of *Cry j* I and a sixteen amino acid internal sequence have been determined (Taniai supra).

A second allergen has recently been isolated from the pollen of Cryptomeria japonica (Japanese cedar) (Sakaguchi et al. (1990) *Allergy* 45:309-312). This allergen, designated *Cry j* II, has been reported to have a molecular weight of approximately 37 kDa and 45 kDa when assayed on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under non-reducing and reducing conditions, respectively (Sukaguchi et al., supra). *Cry j* II was found to have no immunological cross-reactivity with *Cry j* I (Sakaguchi (1990) supra; Kawashima et al. (1992) *Int. Arch. Allergy Immunol.* 98:110-117). Most patients with Japanese cedar pollinosis were found to have IgE antibodies to both *Cry j* I and *Cry j* II. however, 29% of allergic patients had IgE that only reacted with *Cry j* I and 14% of allergic patients had IgE that only reacted with *Cry j* II (Sakaguchi (1990) supra). Isoelectric focusing of *Cry j* II indicated that this protein has a pI above 9.5, as compared to pI 8.6-8.8 for *Cry j* I (Sakaguchi (1990) supra). Further, the reported NH_2-terminal sequence for *Cry j* II, NH_2-AlaIleAsnIlePheAsnValGluLysTyr-COOH, did not match that reported for *Cry j* I (Sakaguchi (1990) supra).

Despite the attention Japanese cedar pollinosis allergens have received, definition or characterization of the allergens responsible for its adverse effects on people is far from complete. Current desensitization therapy involves treatment with pollen extract with its attendant risks of anaphylaxis if high doses of pollen extract are administered, or long desensitization times when low doses of pollen extract are administered.

Summary of the Invention

The present invention provides nucleic acid sequences coding for the *Cryptomeria japonica* major pollen allergen *Cry j* II and fragments thereof. The present invention also provides purified *Cry j* II and at least one fragment thereof

produced in a host cell transformed with a nucleic acid sequence coding for *Cry j* II or at least one fragment thereof and fragments of *Cry j* II prepared synthetically. As used herein, a fragment of the nucleic acid sequence coding for the entire amino acid sequence of *Cry j* II refers to a nucleotide sequence having fewer bases than the nucleotide sequence coding for the entire amino acid sequence of *Cry j* II and/or mature *Cry j* II. *Cry j* II and fragments thereof are useful for diagnosing, treating, and preventing Japanese cedar pollinosis. This invention is more particularly described in the appended claims and is described in its preferred embodiments in the following description.

Description of the Drawings

Fig. 1a shows an SDS-PAGE (12%) analysis of *Cry j* II under non-reducing conditions.

Fig. 1b shows an SDS-PAGE (12%) analysis of *Cry j* II under reducing conditions.

Fig. 2 shows the results of mono S column chromatography of *Cry j* II eluted with a step gradient of NaCl in 10mM sodium acetate buffer, pH 5.0.

Fig. 3 shows an SDS-PAGE (12%) of purified subfractions of *Cry j* II analyzed under reducing conditions.

Fig. 4 shows the nucleic acid sequence (SEQ ID NO: 1) and the deduced amino acid (SEQ ID NO: 2) coding for *Cry j* II.

Fig. 5 shows the deduced amino acid sequence of *Cry j* II (SEQ ID NO: 2).

Fig. 6 shows the long form (SEQ ID NO: 4) and short form (SEQ ID NO: 5) NH_2-terminii amino acid sequences of *Cry j* II determined by protein sequence analysis as discussed in Example 2 aligned with the ten amino acid sequence of *Cry j* II (SEQ ID NO: 3) defined by Sakaguchi et al., supra (SEQ ID NO: 6).

Fig. 7 is a graphic representation of the results of a direct ELISA assay showing the binding response of the monoclonal antibody 4B11 and seven patients' (Batch 1) plasma IgE to purified *Cry j* I as the coating antigen.

Fig. 8 is a graphic representation of a direct ELISA assay showing the binding response of the monoclonal antibody 4B11, and seven patients' (Batch 1) plasma IgE to purified native *Cry j* II as the coating antigen.

Fig. 9 is a graphic representation of a direct ELISA assay showing the binding response of the monoclonal antibody, 4B11, and seven patients' (Batch 1) plasma IgE to recombinant *Cry j* II (r*Cry j* II) as the coating antigen.

Fig. 10 is a graphic representation of a direct ELISA assay showing the binding response of eight patients' (Batch 2) plasma IgE to purified native *Cry j* I.

Fig. 11 is a graphic representation of a direct ELISA assay showing the binding response of eight patients' (Batch 2) plasma IgE to purified native *Cry j* II.

Fig. 12 is a graphic representation of a direct ELISA assay showing the binding response of eight patients' (Batch 2) plasma IgE to recombinant *Cry j* II.

Fig. 13 is a graphic representation of a direct ELISA assay showing the binding response of eight patients' (Batch 3) plasma IgE to purified native *Cry j* I.

Fig. 14 is a graphic representation of a direct ELISA assay showing the binding response of eight patients' (Batch 3) plasma IgE to purified native *Cry j* II.

Fig. 15 is a graphic representation of a direct ELISA assay showing the binding response of eight patients' (Batch 3) plasma IgE to recombinant *Cry j* II.

Fig. 16 is a table which summarizes both the MAST scores performed on patient's plasma samples (Batch 1-3) and the direct ELISA results shown in Figs. 7-15; a positive response is indicated by a (+) sign and the number of positive responses for each antigen is shown at the bottom of each column.

Detailed Description of the Invention

The present invention provides nucleic acid sequences coding for *Cry j* II, an allergen found in Japanese cedar pollen. The nucleic acid sequence coding for *Cry j* II shown in Fig. 4 (SEQ ID NO: 1) encodes a protein of 514 amino acids. The deduced *Cry j* II amino acid sequence is shown in Figs. 4 and 5 (SEQ ID NO: 2) . Direct protein sequence analysis of native purified *Cry j* II resulted in two separate overlapping NH_2-termini sequences, designated Long and Short, corresponding respectively to amino acids 46 through 89 (SEQ ID NO: 4) and 51 through 89 (SEQ ID NO: 5) of Figs. 4, 5 and 6. The ten amino acid sequence NH_2-AlaIleAsnIlePhe-AsnValGluLysTry-COOH (SEQ ID NO: 6) previously defined by Sakaguchi et al, supra for *Cry j* II corresponds to amino acids 55 through 64 of Figs. 4 and 6. The full-length *Cry j* II sequence contains 20 cysteine residues and three potential N-linked glycosylation sites with the consensus sequence of Asn-Xxx-Ser/Thr. According to the program contained in *PC Gene, Intelligenetics* (Mountain View, CA) the proteins with the NH_2-termini defined by the Long and Short forms of *Cry j* II would contain 469 and 464 amino acids, respectively, and have predicted molecular weights of 51.5 kDa (long) and 50.9 kDa (short). The amino acid sequence representing the long form of *Cry j* II is encoded by the nucleotide sequence extending from bases 177-1586 (SEQ ID NO: 7) as shown in Fig. 4, and the amino acid sequence representing the short form of *Cry j* II is encoded by the nucleotide sequence extending from 192-1586 (SEQ ID NO: 8) as shown in Fig. 4. A host cell transformed with a vector containing the cDNA insert coding for full-

length *Cry j* II has been deposited with the American Type Culture Collection, ATCC No. 69105.

Fragments of the nucleic acid sequence coding for fragments of *Cry j* II are also within the scope of the invention. Fragments within the scope of the invention include those coding for parts of *Cry j* II which induce an immune response in mammals, preferably humans, such as stimulation of minimal amounts of IgE; binding of IgE; eliciting the production of IgG and IgM antibodies; or the eliciting of a T cell response such as proliferation and/or lymphokine secretion and/or the induction of T cell anergy. The foregoing fragments of *Cry j* II are referred to herein as antigenic fragments. Fragments within the scope of the invention also include those capable of hybridizing with nucleic acid from other plant species for use in screening protocols to detect allergens that are cross-reactive with *Cry j* II. As used herein, a fragment of the nucleic acid sequence coding for *Cry j* II refers to a nucleotide sequence having fewer bases than the nucleotide sequence coding for the entire amino acid sequence of *Cry j* II and/or mature *Cry j* II. Generally, the nucleic acid sequence coding for the fragment or fragments of *Cry j* II will be selected from the bases coding for the mature protein, however, in some instances it may be desirable to select all or a part of a fragment or fragments from the leader sequence portion of the nucleic acid sequence of the invention. The nucleic acid sequence of the invention may also contain linker sequences, modified restriction endonuclease sites and other sequences useful for cloning, expression or purification of *Cry j* II or fragments thereof.

A nucleic acid sequence coding for *Cry j* II may be obtained from *Cryptomeria japonica* plants. Applicants have found that fresh pollen and staminate cones are a good source of *Cry j* II mRNA. It may also be possible to obtain the nucleic acid sequence coding for *Cry j* II from genomic DNA. *Cryptomeria japonica* is a well-known species of cedar, and plant material may be obtained from wild, cultivated, or ornamental plants. The nucleic acid sequence coding for *Cry j* II may be obtained using the method disclosed herein or any other suitable techniques for isolation and cloning of genes. The nucleic acid sequence of the invention may be DNA or RNA.

The present invention provides expression vectors and host cells transformed to express the nucleic acid sequences of the invention. Nucleic acid coding for *Cry j* II, or at least one fragment thereof may be expressed in bacterial cells such as *E. coli*, insect cells (baculovirus), yeast, or mammalian cells such as Chinese hamster ovary cells (CHO). Suitable expression vectors, promoters, enhancers, and other expression control elements may be found in Sambrook et al. *Molecular Cloning: A*

Laboratory Manual, second edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). Other suitable expression vectors, promoters, enhancers, and other expression elements are known to those skilled in the art. Expression in mammalian, yeast or insect cells leads to partial or complete glycosylation of the recombinant material and formation of any inter- or intra-chain disulfide bonds. Suitable vectors for expression in yeast include YepSec1 (Baldari et al. (1987) *Embo J.* 6: 229-234); pMFa (Kurjan and Herskowitz (1982) *Cell* 30: 933-943); JRY88 (Schultz et al. (1987) *Gene* 54: 113-123) and pYES2 (Invitrogen Corporation, San Diego, CA). These vectors are freely available. Baculovirus and mammalian expression systems are also available. For example, a baculovirus system is commercially available (PharMingen, San Diego, CA) for expression in insect cells while the pMSG vector is commerically available (Pharmacia, Piscataway, NJ) for expression in mammalian cells.

For expression in *E. coli*, suitable expression vectors include, among others, pTRC (Amann et al. (1988) *Gene* 69: 301-315); pGEX (Amrad Corp., Melbourne, Australia); pMAL (N.E. Biolabs, Beverly, MA); pRIT5 (Pharmacia, Piscataway, NJ); pET-11d (Novagen, Madison, WI) Jameel et al., (1990) *J. Virol.* 64:3963-3966; and pSEM (Knapp et al. (1990) *BioTechniques* 8: 280-281). The use of pTRC, and pET-11d, for example, will lead to the expression of unfused protein. The use of pMAL, pRIT5 pSEM and pGEX will lead to the expression of allergen fused to maltose E binding protein (pMAL), protein A (pRIT5), truncated ß-galactosidase (PSEM), or glutathione S-transferase (pGEX). When *Cry j* II, fragment, or fragments thereof is expressed as a fusion protein, it is particularly advantageous to introduce an enzymatic cleavage site at the fusion junction between the carrier protein and *Cry j* II or fragment thereof. *Cry j* II or fragment thereof may then be recovered from the fusion protein through enzymatic cleavage at the enzymatic site and biochemical purification using conventional techniques for purification of proteins and peptides. Suitable enzymatic cleavage sites include those for blood clotting Factor Xa or thrombin for which the appropriate enzymes and protocols for cleavage are commercially available from for example Sigma Chemical Company, St. Louis, MO and N.E. Biolabs, Beverly, MA. The different vectors also have different promoter regions allowing constitutive or inducible expression with, for example, IPTG induction (PRTC, Amann et al., (1988) supra; pET-11d, Novagen, Madison, WI) or temperature induction (pRIT5, Pharmacia, Piscataway, NJ) . It may also be appropriate to express recombinant *Cry j* II in different *E. coli* hosts that have an altered capacity to degrade recombinantly expressed proteins (e.g. U.S. patent 4,758,512). Alternatively, it may be advantageous to alter the nucleic

acid sequence to use codons preferentially utilized by *E. coli*, where such nucleic acid alteration would not affect the amino acid sequence of the expressed protein.

Host cells can be transformed to express the nucleic acid sequences of the invention using conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, or electroporation. Suitable methods for transforming the host cells may be found in Sambrook et al. supra, and other laboratory textbooks. The nucleic acid sequences of the invention may also be synthesized using standard techniques.

The present invention also provides a method of producing purified Japanese cedar pollen allergen *Cry j* II or at least one fragment thereof comprising the steps of culturing a host cell transformed with a DNA sequence encoding Japanese cedar pollen allergen *Cry j* II or at least one fragment thereof in an appropriate medium to produce a mixture of cells and medium containing said Japanese cedar pollen allergen *Cry j* II or at least one fragment thereof; and purifying the mixture to produce substantially pure Japanese cedar pollen allergen *Cry j* II or at least one fragment thereof. Host cells transformed with an expression vector containing DNA coding for *Cry j* II or at least one fragment thereof are cultured in a suitable medium for the host cell. *Cry j* II protein and peptides can be purified from cell culture medium, host cells, or both using techniques known in the art for purifying peptides and proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis and immunopurification with antibodies specific for *Cry j* II or fragments thereof. The terms isolated and purified are used interchangeably herein and refer to peptides, protein, protein fragments, and nucleic acid sequences substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical precursors.

Cry j II protein may also be isolated from Japanese cedar pollen as described in Example 1. *Cry j* II isolated directly from Japanese cedar pollen is referred to herein as "purified native" *Cry j* II. It is preferable that purified native *Cry j* II of the invention be at least 80% pure, and more preferably at least 90% pure and even more preferably be purified to homogeneity (at least 99% pure).

Another aspect of the invention provides preparations comprising Japanese cedar pollen allergen *Cry j* II or at least one fragment thereof synthesized in a host cell transformed with a DNA sequence encoding all or a portion of Japanese cedar pollen allergen *Cry j* II, or chemically synthesized, and purified Japanese cedar pollen allergen *Cry j* II protein, or at least one antigenic fragment thereof produced in a host cell transformed with a nucleic acid sequence of the invention, or chemically synthesized. In preferred embodiments of the invention the *Cry j* II

protein is produced in a host cell transformed with the nucleic acid sequence coding for at least the mature *Cry j* II protein.

Fragments of an allergen from *Cry j* II, eliciting a desired antigenic response (referred to herein as antigenic fragments) are defined herein as any protein fragment or peptide which can be derived from the *Cry j* II proteins, but does not include the ten amino acid fragments which extends from amino acid residues 55-64, as shown in Figs. 4, 5 and 6, but may include any portion of that ten amino acid fragment in conjunction with another fragment derived from *Cry j* II. Antigenic fragments of *Cry j* II may be obtained, for example, by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid sequence of the invention coding for such peptides, or by screening peptides which have been synthesized chemically using techniques known in the art, or by screening peptides produced by chemical cleavage of the allergen. The allergen may be arbitrarily divided into fragments of a desired length with no overlap of the peptides, or preferably divided into fragments of a desired length with no overlap of the peptides, or preferably divided into overlapping fragments of a desired length. The fragments are tested to determine their antigenicity (e.g. the ability of the fragment to induce an immune response such as T cell proliferation as discussed in Example 7).

Antigenic fragments may also be predicted using an algorithm such as that discussed in a paper by Hill et al, *Journal of Immunology,* 147:184-197 (1991). Algorithms for predicting peptides which elicit T cell activity such as the algorithm discussed by Hill et al. are based on the protein's sequence wherein certain patterns within the sequence are likely to bind MHC and therefore may contain T cell epitopes. The peptides predicted by the algorithm such as Cry j IIA and Cry j IIB discussed in Example 7 may be produced recombinantly or synthetically and tested for T cell activity as discussed in Example 7.

If fragments of Japanese cedar pollen allergen, e.g. *Cry j* II are to be used for therapeutic purposes, then the fragments of Japanese cedar pollen allergen which are capable of eliciting a T cell response such as stimulation (i.e., proliferation or lymphokine secretion) and/or are capable of inducing T cell anergy are particularly desirable and fragments of Japanese cedar pollen which have minimal IgE stimulating activity are also desirable. Additionally, for therapeutic purposes, purified Japanese cedar pollen allergens, e.g. *Cry j* II, and fragments thereof preferably do not bind IgE specific for Japanese cedar pollen or bind such IgE to a substantially lesser extent than the purified native Japanese cedar pollen allergen binds such IgE. If the purified Japanese cedar pollen allergen or fragment or fragments thereof bind IgE, it is preferable that such binding does not result in the

release of mediators (e.g. histamines) from mast cells or basophils. Minimal IgE stimulating activity refers to IgE stimulating activity that is less than the amount of IgE production stimulated by the native *Cry j* II protein.

Isolated antigenic fragments or peptides of the present invention which have T cell stimulating activity, and thus comprise at least one T cell epitope are particularly desirable. T cell epitopes are believed to be involved in initiation and perpetuation of the immune response to a protein allergen which is responsible for the clinical symptoms of allergy. These T cell epitopes are thought to trigger early events at the level of the T helper cell by binding to an appropriate HLA molecule on the surface of an antigen presenting cell and stimulating the relevant T cell subpopulation. These events lead to T cell proliferation, lymphokine secretion, local inflammatory reactions, recruitment of additional immune cells to the site, and activation of the B cell cascade leading to production of antibodies. One isotype of these antibodies, IgE, is fundamentally important to the development of allergic symptoms and its production is influenced early in the cascade of events, at the level of the T helper cell, by the nature of the lymphokines secreted. An epitope is the basic element or smallest unit of recognition by a receptor, particularly immunoglobulins, histocompatibility antigens and T cell receptors, where the epitope comprises amino acids essential to receptor recognition. Amino acid sequences which mimic those of the epitopes particularly T cell epitopes and which modify the allergic response to protein allergens including those capable of down regulating allergic response to Cry j II, are within the scope of this invention.

As discussed in Example 7, human T cell stimulating activity can be tested by culturing T cells obtained from an individual sensitive to Japanese cedar pollen allergen, (i.e., an individual who has an IgE mediated immune response to Japanese cedar pollen allergen) with a peptide derived from the allergen and determining whether proliferation of T cells occurs in response to the peptide as measured, e.g., by cellular uptake of tritiated thymidine. Stimulation indices for responses by T cells to peptides can be calculated as the maximum CPM in response to a peptide divided by the control CPM. A stimulation index (S.I.) equal to or greater than two times the background level is considered "positive". Positive results are used to calculate the mean stimulation index for each peptide tested. Preferred peptides of this invention comprise at least one T cell epitope and have a mean T cell stimulation index of greater than or equal to 2.0. A peptide having a mean T cell stimulation index of greater than or equal to 2.0 is considered useful as a therapeutic agent. As shown in Fig. 17 *Cry j* II peptides Cry j IIA and Cry j IIB have mean stimulation indexes of at least two and therefore comprise at least one T cell epitope as

predicted.

Purified protein allergens from Japanese cedar pollen or preferred antigenic fragments thereof, when administered to a Japanese cedar pollen-sensitive individual, or an individual allergic to an allergen cross-reactive with Japanese cedar pollen allergen, are capable of modifying the allergic response of the individual to Japanese cedar pollen or such cross-reactive allergen of the individual, and preferably are capable of modifying the B-cell response, T-cell response or both the B-cell and the T-cell response of the individual to the allergen. As used herein, modification of the allergic response of an individual sensitive to a Japanese cedar pollen allergen can be defined as non-responsiveness or diminution in symptoms to the allergen, as determined by standard clinical procedures (See e.g. Varney et al, *British Medical Journal,* 302:265-269 (1990)) including diminution in Japanese cedar pollen induced asthmatic symptoms. As referred to herein, a diminution in symptoms includes any reduction in allergic response of an individual to the allergen after the individual has completed a treatment regimen with a peptide or protein of the invention. This diminution may be subjective (i.e. the patient feels more comfortable in the presence of the allergen). Diminution in symptoms can be determined clinically as well, using standard skin tests as is known in the art.

The purified *Cry j* II protein or fragments thereof are preferably tested in mammalian models of Japanese cedar pollinosis such as the mouse model disclosed in Tamura et al. (1986) *Microbiol. Immunol.* 30: 883-896, or U.S. patent 4,939,239; or the primate model disclosed in Chiba et al. (1990) *Int. Arch. Allergy Immunol.* 93: 83-88. Initial screening for IgE binding to the protein or fragments thereof may be performed by scratch tests or intradermal skin tests on laboratory animals or human volunteers, or in *in vitro* systems such as RAST (radioallergosorbent test), RAST inhibition, ELISA assay, radioimmunoassay (RIA), or histamine release.

Exposure of allergic individuals to purified protein allergens of the present invention or to the antigenic fragments of the present invention which comprise at least one T cell epitope and are derived from protein allergens may tolerize or anergize appropriate T cell subpopulations such that they become unresponsive to the protein allergen and do not participate in stimulating an immune response upon such exposure. In addition, administration of the protein allergen of the invention or an antigenic fragment of the present invention which comprises at least one T cell epitope may modify the lymphokine secretion profile as compared with exposure to the naturally-occurring protein allergen or portion thereof (e.g. result in a decrease of IL-4 and/or an increase in IL-2). Furthermore, exposure to such antigenic

fragment or protein allergen may influence T cell subpopulations which normally participate in the response to the allergen such that these T cells are drawn away from the site(s) of normal exposure to the allergen (e.g., nasal mucosa, skin, and lung) towards the site(s) of therapeutic administration of the fragment or protein allergen. This redistribution of T cell subpopulations may ameliorate or reduce the ability of an individual's immune system to stimulate the usual immune response at the site of normal exposure to the allergen, resulting in a dimunution in allergic symptoms.

The isolated *Cry j* II protein, and fragments or portions derived therefrom can be used in methods of diagnosing, treating and preventing allergic reactions to Japanese cedar pollen allergen or a cross reactive protein allergen. Thus the present invention provides therapeutic compositions comprising purified Japanese cedar pollen allergen *Cry j* II or at least one fragment thereof produced in a host cell transformed to express *Cry j* II or at least one fragment thereof, and a pharmaceutically acceptable carrier or diluent. The therapeutic compositions of the invention may also comprise synthetically prepared *Cry j* II or at least one fragment thereof and a pharmaceutically acceptable carrier or diluent. Administration of the therapeutic compositions of the present invention to an individual to be desensitized can be carried out using known techniques. *Cry j* II protein or at least one fragment thereof may be administered to an individual in combination with, for example, an appropriate diluent, a carrier and/or an adjuvant. Pharmaceutically acceptable diluents include saline and aqueous buffer solutions. Pharmaceutically acceptable carriers include polyethylene glycol (Wie et al. (1981) *Int. Arch. Allergy Appl. Immunol.* 64:84-99) and liposomes (Strejan et al. (1984) *J. Neuroimmunol* 7: 27). For purposes of inducing T cell anergy, the therapeutic composition is preferably administered in nonimmunogenic form, e.g. it does not contain adjuvant. Such compositions will generally be administered by injection (subcutaneous, intravenous, etc.), oral administration, inhalation, transdermal application or rectal administration. The therapeutic compositions of the invention are administered to Japanese cedar pollen-sensitive individuals at dosages and for lengths of time effective to reduce sensitivity (i.e, reduce the allergic response) of the individual to Japanese cedar pollen. Effective amounts of the therapeutic compositions will vary according to factors such as the degree of sensitivity of the individual to Japanese cedar pollen, the age, sex, and weight of the individual, and the ability of the *Cry j* II protein or fragment thereof to elicit an antigenic response in the individual.

The *Cry j* II cDNA (or the mRNA from which it was transcribed) or a portion thereof can be used to identify similar sequences in any variety or type of

plant and thus, to identify or "pull out" sequences which have sufficient homology to hybridize to the *Cry j* II cDNA or mRNA or portion thereof, for example, DNA from allergens of *Cupressus sempervirens, Juniperus sabinoides etc.*, under conditions of low stringency. Those sequences which have sufficient homology (generally greater than 40%) can be selected for further assessment using the method described herein. Alternatively, high stringency conditions can be used. In this manner, DNA of the present invention can be used to identify, in other types of plants, preferably related families, genera, or species such as *Juniperus*, or *Cupressus*, sequences encoding polypeptides having amino acid sequences similar to that of Japanese cedar pollen allergen *Cry j* II, and thus to identify allergens in other species. Thus, the present invention includes not only *Cry j* II, but also other allergens encoded by DNA which hybridizes to DNA of the present invention. The invention further includes previously unidentified isolated allergenic proteins or fragments thereof that are immunologically related to *Cry j* II or fragments thereof, such as by antibody cross-reactivity wherein the isolated allergenic proteins or fragments thereof are capable of binding to antibodies specific for the protein and peptides of the invention, or by T cell cross-reactivity wherein the isolated allergenic proteins or fragments thereof are capable of stimulating T cells specific for the protein and peptides of this invention.

Proteins or peptides encoded by the cDNA of the present invention can be used, for example as "purified" allergens. Such purified allergens are useful in the standardization of allergen extracts which are key reagents for the diagnosis and treatment of Japanese cedar pollinosis. Furthermore, by using peptides based on the nucleic acid sequences of *Cry j* II, anti-peptide antisera or monoclonal antibodies can be made using standard methods. These sera or monoclonal antibodies can be used to standardize allergen extracts.

Through use of the peptides and protein of the present invention, preparations of consistent, well-defined composition and biological activity can be made and administered for therapeutic purposes (e.g. to modify the allergic response of a Japanese cedar sensitive individual to pollen of such trees). Administration of such peptides or protein may, for example, modify B-cell response to *Cry j* II allergen, modify T-cell response to *Cry j* II allergen or modify both B-cell and T-cell responses. Purified peptides can also be used to study the mechanism of immunotherapy of *Cryptomeria japonica* allergy and to design modified derivatives or analogues useful in immunotherapy.

Work by others has shown that high doses of allergens generally produce the best results (i.e., best symptom relief). However, many people are unable to

tolerate large doses of allergens because of allergic reactions to the allergens. Modification of naturally-occurring allergens can be designed in such a manner that modified peptides or modified allergens which have the same or enhanced therapeutic properties as the corresponding naturally-occurring allergen but have reduced side effects (especially anaphylactic reactions) can be produced. These can be, for example, a protein or peptide of the present invention (e.g., one having all or a portion of the amino acid sequence of *Cry j* II), or a modified protein or peptide, or protein or peptide analogue.

It is possible to modify the structure of a protein or peptide of the invention for such purposes as increasing solubility, enhancing therapeutic or preventive efficacy, or stability (e.g., shelf life ex vivo, and resistance to proteolytic degradation in vivo). A modified protein or peptide can be produced in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition, to modify immunogenicity and/or reduce allergenicity, or to which a component has been added for the same purpose. For example, the amino acid residues essential to T cell epitope function can be determined using known techniques (e.g., substitution of each residue and determination of the presence or absence of T cell reactivity).

For example, a peptide can be modified so that it maintains the ability to induce T cell anergy and bind MHC proteins without the ability to induce a strong proliferative response or possibly any proliferative response when administered in immunogenic form. In this instance, critical binding residues for the T cell receptor can be determined using known techniques (e.g., substitution of each residue and determination of the presence or absence of T cell reactivity). Those residues shown to be essential to interact with the T cell receptor can be modified by replacing the essential amino acid with another, preferably similar amino acid residue (a conservative substitution) whose presence is shown to enhance, diminish but not eliminate binding to relevant MHC.

Additionally, peptides of the invention can be modified by replacing an amino acid shown to be essential to interact with the MHC protein complex with another, preferably similar amino acid residue (conservative substitution) whose presence is shown to enhance, diminish but not eliminate or not effect T cell activity. In addition, amino acid residues which are not essential for interaction with the MHC protein complex but which still bind the MHC protein complex can be modified by being replaced by another amino acid whose incorporation may enhance, not effect, or diminish but not eliminate T cell reactivity. Preferred amino acid substitutions for non-essential amino acids include, but are not limited to

substitutions with alanine, glutamic acid, or a methyl amino acid.

Another example of a modification of protein or peptides is substitution of cysteine residues preferably with alanine, serine, threonine, leucine or glutamic acid to minimize dimerization via disulfide linkages. Another example of modification of the peptides of the invention is by chemical modification of amino acid side chains or cyclization of the peptide.

In order to enhance stability and/or reactivity, the protein or peptides of the invention can also be modified to incorporate one or more polymorphisms in the amino acid sequence of the protein allergen resulting from natural allelic variation. Additionally, D-amino acids, non-natural amino acids or non-amino acid analogues can be substituted or added to produce a modified protein or peptide within the scope of this invention. Furthermore, proteins or peptides of the present invention can be modified using the polyethylene glycol (PEG) method of A. Sehon and co-workers (Wie et al. supra) to produce a protein or peptide conjugated with PEG. In addition, PEG can be added during chemical synthesis of a protein or peptide of the invention. Modifications of proteins or peptides or portions thereof can also include reduction/ alyklation (Tarr in: *Methods of Protein Microcharacterization,* J.E. Silver ed. Humana Press, Clifton, NJ, pp 155-194 (1986)); acylation (Tarr, supra); chemical coupling to an appropriate carrier (Mishell and Shiigi, eds, *Selected Methods in Cellular Immunology,* WH Freeman, San Francisco, CA (1980); U.S. Patent 4,939,239; or mild formalin treatment (Marsh *International Archives of Allergy and Applied Immunology,* 41:199-215 (1971)).

To facilitate purification and potentially increase solubility of proteins or peptides of the invention, it is possible to add reporter group(s) to the peptide backbone. For example, poly-histidine can be added to a peptide to purify the peptide on immobilized metal ion affinity chromatography (Hochuli, E. et al., *Bio/Technology,* 6:1321-1325 (1988)). In addition, specific endoprotease cleavage sites can be introduced, if desired, between a reporter group and amino acid sequences of a peptide to facilitate isolation of peptides free of irrelevant sequences. In order to successfully desensitize an individual to a protein antigen, it may be necessary to increase the solubility of a protein or peptide by adding functional groups to the peptide or by not including hydrophobic T cell epitopes or regions containing hydrophobic epitopes in the peptides or hydrophobic regions of the protein or peptide.

To potentially aid proper antigen processing of T cell epitopes within a peptide, canonical protease sensitive sites can be recombinantly or synthetically engineered between regions, each comprising at least one T cell epitope. For

example, charged amino acid pairs, such as KK or RR, can be introduced between regions within a peptide during recombinant construction of the peptide. The resulting peptide can be rendered sensitive to cathepsin and/or other trypsin-like enzymes cleavage to generate portions of the peptide containing one or more T cell epitopes. In addition, such charged amino acid residues can result in an increase in solubility of a peptide.

Site-directed mutagenesis of DNA encoding a peptide or protein of the invention (e.g. *Cry j* II or a fragment thereof) can be used to modify the structure of the peptide or protein by methods known in the art. Such methods may, among others, include PCR with degenerate oligonucleotides (Ho et al., *Gene,* 77:51-59 (1989)) or total synthesis of mutated genes (Hostomsky, Z. et al., *Biochem. Biophys, Res. Comm.,*161:1056-1063 (1989)). To enhance bacterial expression, the aforementioned methods can be used in conjunction with other procedures to change the eucaryotic codons in DNA constructs encoding protein or peptides of the invention to ones preferentially used in *E. coli*, yeast, mammalian cells, or other eukaryotic cells.

Using the structural information now available, it is possible to design *Cry j* II peptides which, when administered to a Japanese cedar pollen sensitive individual in sufficient quantities, will modify the individual's allergic response to Japanese cedar pollen. This can be done, for example, by examining the structure of *Cry j* II, producing peptides (via an expression system, synthetically or otherwise) to be examined for their ability to influence B-cell and/or T-cell responses in Japanese cedar pollen sensitive individuals and selecting appropriate peptides which contain epitopes recognized by the cells. It is now also possible to design an agent or a drug capable of blocking or inhibiting the ability of Japanese cedar pollen allergen to induce an allergic reaction in Japanese cedar pollen sensitive individuals. Such agents could be designed, for example, in such a manner that they would bind to relevant anti-*Cry j* II IgEs, thus preventing IgE-allergen binding and subsequent mast cell degranulation. Alternatively, such agents could bind to cellular components of the immune system, resulting in suppression or desensitization of the allergic response to *Cryptomeria japonica* pollen allergens. A non-restrictive example of this is the use of appropriate B- and T-cell epitope peptides, or modifications thereof, based on the cDNA/protein structures of the present invention to suppress the allergic response to Japanese cedar pollen. This can be carried out by defining the structures of B- and T-cell epitope peptides which affect B- and T-cell function in *in vitro* studies with blood components from Japanese cedar pollen sensitive individuals.

Protein, peptides or antibodies of the present invention can also be used for detecting and diagnosing Japanese cedar pollinosis. For example, this could be done by combining blood or blood products obtained from an individual to be assessed for sensitivity to Japanese cedar pollen with an isolated antigenic peptide or peptides of *Cry j* II, or isolated *Cry j* II protein, under conditions appropriate for binding of components in the blood (e.g., antibodies, T-cells, B-cells) with the peptide(s) or protein and determining the extent to which such binding occurs. Other diagnostic methods for allergic diseases which the protein, peptides or antibodies of the present invention can be used include radio-allergosorbent test (RAST), paper radioimmunosorbent test (PRIST), enzyme linked immunosorbent assay (ELISA), radioimmunoassays (RIA), immuno-radiometric assays (IRMA), luminescence immunoassays (LIA), histamine release assays and IgE immunoblots.

In another diagnostic test, the presence in individuals of IgE specific for *Cry j II* at least one protein allergen and the ability of T cells of the individuals to respond to T cell epitope(s) of *Cry j II* protein allergen can be determined by administering to the individuals an Immediate Type Hypersensitivity test and a Delayed Type Hypersensitivity test. The individuals are administered an Immediate Type Hypersensitivity test (see e.g. *Immunology* (1985) Roitt, I.M., Brostoff, J., Male, D.K. (eds), C.V. Mosby Co., Gower Medical Publishing, London, NY, pp. 19.2-19.18; pp. 22.1-22.10) utilizing the *Cry i II* protein allergen or a portion thereof, or a modified form of the *Cry j II* protein allergen or a portion thereof, each of which binds IgE specific for the allergen. The same individuals are administered a Delayed Type Hypersensitivity test prior to, simultaneously with, or subsequent to administraiton of the Immediate Type Hypersensitivity test. Of course, if the Immediate Type Hypersensitivity test is administered prior to the Delayed Type Hypersensitivity test, the Delayed Type Hypersensitivity test would be given to those individuals exhibiting a specific Immediate Type Hypersensitivity reaction. The Delayed Type Hypersensitivity test utilizes a modified form of the protein allergen or a portion thereof, the protein allergen produced recombinantly, or a recombitope peptide derived from the protein allergen, each of which has human T cell stimulating activity and each of which does not bind IgE specific for the allergen in a substantial percentage of the population of individuals sensitive to the allergen (e.g., at least about 75%). Based on the results of the above diagnostic tests, those individuals found to have both a specific Immediate Type Hypersensitivity reaction and a specific Delayed Type Hypersensitivity reaction are suitable candidates for administration of a therapeutically effective amount of a therapeutic composition. The therapeutic composition comprises the modified form of the protein or portion

thereof, the recombinantly produced protein allergen, or the recombitope peptide, each as used in the Delayed Type Hypersensitivity test, and a pharmaceutically acceptable carrier or diluent.

The present invention also provides a method of producing *Cry j* II or fragment thereof comprising culturing a host cell containing an expression vector which contains DNA encoding all or at least one fragment of *Cry j* II under conditions appropriate for expression of *Cry j* II or at least one fragment. The expressed product is then recovered, using known techniques. Alternatively, *Cry j* II or fragment thereof can be synthesized using known mechanical or chemical techniques.

The DNA used in any embodiment of this invention can be cDNA obtained as described herein, or alternatively, can be any oligodeoxynucleotide sequence having all or a portion of a sequence represented herein, or their functional equivalents. Such oligodeoxynucleotide sequences can be produced chemically or enzymatically, using known techniques. A functional equivalent of an oligonucleotide sequence is one which is 1) a sequence capable of hybridizing to a complementary oligonucleotide to which the sequence (or corresponding sequence portions) of *Cry j* II or fragments thereof hybridizes, or 2) the sequence (or corresponding sequence portion) complementary to *Cry j* II, and/or 3) a sequence which encodes a product (e.g., a polypeptide or peptide) having the same functional characteristics of the product encoded by the sequence (or corresponding sequence portion) of *Cry j* II. Whether a functional equivalent must meet one or both criteria will depend on its use (e.g., if it is to be used only as an oligoprobe, it need meet only the first or second criteria and if it is to be used to produce a *Cry j* II allergen, it need only meet the third criterion).

The invention is further illustrated by the following non-limiting examples.

Example 1

Purification of Native Japanese Cedar Pollen Allergen (*Cry j* II)

The following purification of native *Cry j* II from Japanese cedar pollen was modified from previously published reports (Yasueda et al, *J. Allergy Clin. Immunol.* 71:77 (1983); Sukaguchi et al., *Allergy*, 45:309 (1990)).

100g of Japanese cedar pollen obtained from Japan (Hollister-Stier, Spokane, WA) was defatted in 1L diethyl ether three times, the pollen was collected after filtration and the ether was dried off in a vacuum.

The defatted pollen was extracted at 4°C overnight in 2L extraction buffer

containing 50 mM tris-HCl, pH 7.8, 0.2 M NaCl and protease inhibitors in final concentrations: soybean trypsin inhibitor (2 μg/mL), leupeptin (1 μg/mL), pepstatin A (1 μg/mL) and phenyl methyl sulfonyl fluoride (0.17 mg/mL). The insoluble material was re-extrated with 1.2L extraction buffer at 4°C overnight and both extracts were combined together and depigmented by batch absorption with Whatman DE-52 (200g dry weight) equilibrated with the extraction buffer.

The depigmented material was then fractionated by ammonium sulfate precipitation at 80% saturation (4°C), which removed much of the lower molecular weight material. The resulting pellet was resuspended in 0.4 L of 50 mM Na-acetate, pH 5.0 containing protease inhibitors and was dialyzed extensively against the same buffer.

The sample was further subjected to purification by either one of the two methods described below.

Method A

The sample was applied to a 100 mL DEAE cellulose column (Whatman DE-52) equilibrated at 4°C with 50 mM Na-acetate, pH 5.0 with protease inhibitors. The unbound material (basic proteins) from the DEAE cellulose column was then applied to a 50 ml cation exchange column (Whatman CM-52) which was equilibrated with 10 mM Na-acetate, pH 5.0 at 4°C with protease inhibitors. A linear gradient of 0-0.3 M NaCl was used to elute the proteins. The early fractions were enriched in *Cry j* I whereas the later fractions were enriched in *Cry j* II. Fractions containing *Cry j* II were pooled and next applied to an 1 mL Mono S HR 5/5 column (Pharmacia, Piscataway, NJ) in 10 mM Na-acetate, pH 5.0, and proteins were eluted with a linear gradient of NaCl at room temperature. Residual *Cry j* I was eluted at ~0.2 M NaCl and *Cry j* II was eluted between 0.3 to 0.4 M NaCl. The *Cry j* II peak was pooled and concentrated to twofold by lyophilization and subjected to gel filtration chromatography.

The sample was applied to FPLC Superdex 75 16/60 column (Pharmacia, Piscataway, NJ) in 10 mM acetate buffer, pH 5.0 and 0.15 M NaCl at a flow rate of 30 ml/min. at room temperature. Purified *Cry j* II was recovered in the 35-30 kD region. *Cry j* II migrated as two broad bands lower than *Cry j* I under non-reducing conditions (Fig. 1a) but both bands shifted upward and migrated as *Cry j* I under reducing condition (Fig. 1b) when analyzed by silver-stained SDS-PAGE. This highly purified *Cry j* II still contained a small amount (~5%) of *Cry j* I as detected by Western blot using MAb CBF2, which has been shown to bind to *Cry j* I and by N-terminal protein sequencing. This *Cry j* II preparation was used to generate primary

protein sequence of *Cry j* II as described below.

Method B

The dialyzed sample from the ammonium sulfate precipitation was applied at 1 ml/min to an 5.0 ml Q-Sepharose Econapac anion exchange cartridge (BioRad, Richmond, CA) equilibrated with 50 mM Na-acetate, pH 5.0 with protease inhibitors at 4°C. Elution was performed with the above buffer containing 0.5 M NaCl. The basic unbound material was then applied to a 5.0 ml CM-Sepharose Econopac cation exchange cartridge (BioRad, Richmond, CA) equilibrated in 50 mM sodium acetate pH 5.0 with protease inhibitors. Basic proteins were eluted with a linear gradient up to 0.1 M sodium phosphate pH 7.0, 0.3 M NaCl at 1 ml/min at 4°C. A *Cry j* II -enriched peak was collected late in the gradient and further purified by gel filtration chromatography.

FPLC gel filtration was performed using a 320 mL Superdex 75 26/60 (Pharmacia, Piscataway, NJ) column at 0.5 ml/min in 20 mM sodium acetate, pH 5.0, in the presence of 0.15 M NaCl. The major peak containing mostly *Cry j* II eluted between 160 and 190 ml. Contaminating *Cry j* I was next removed by FPLC using a 1.0 ml Mono S 5/5 (Pharmacia, Piscataway, NJ) cation exchange column equilibrated with 10 mM sodium acetate pH 5.0. A stepwise gradient of 0-1 M NaCl was utilized by holding isocratically at 0.2 M, 0.3 M, 0.4 M and 1 M salt concentration.

Multiple peaks (up to nine peaks) were obtained (Fig. 2) and analyzed by silver stained SDS-PAGE under reducing conditions (Fig. 3). *Cry j* I with a reported pI of 8.6-8.9 (Yasueda et al, *J. Allergy Clin. Immunol.*, vol. 17 (1983)), eluted in the earlier peaks and displayed a molecular weight of about 40 kD. *Cry j* II was purified to homogeneity as two bands (Fig. 3) and eluted in the later multiple peaks, suggesting the existence of isoforms. ELISA analysis using the mouse monoclonal 8B11 IgG antibody which was raised against biochemically purified *Cry j* I confirmed the absence of *Cry j* I in these purified *Cry j* II preparation. This purified *Cry j* II was used in the human IgE reactivity studies (Example 6).

Physical properties of *Cry j* II

The physiochemical properties of *Cry j* II were studied and summarized as below. Under non-reducing SDS-PAGE conditions *Cry j* II consists of two bands with molecular weights ranged 34000-32000. The molecular weights of both bands are shifted higher to about 38-36 kD under reducing conditions (Fig. 1b). This shift in SDS-polyacrylamide gel has also been observed by others (Sakaguchi et al,

*Allergy*45:309-312 (1990)). These results suggest that intra-disulfide bonds are probably present in the protein, and it is supported by the present findings that cloned *Cry j* II contains 20 cysteines deduced from the nucleotide sequence (Example 3). The pI of *Cry j* II estimated from IEF gel is about 10. The purified *Cry j* II binds human IgE of some allergic patients.

The two molecular weight bands of *Cry j* II were separated on a 12% SDS-polyacrylamide gel and was then electroblotted onto PVDF membrane (Applied Biosystems, Foster City, CA). The blot was stained with coomassie brilliant blue and was cut and subjected to N-terminal amino acid sequencing. (Example 2). The results showed that the upper and lower molecular weight bands had identical N-terminal sequences except the lower molecular weight band missed the first five amino acids. The estimated molecular weight of the upper band based on the cDNA sequence is about 52,000, which is significantly higher than the molecular weight estimated from SDS-polyacrylamide gel either in the presence or absence of reducing reagent. It is also higher than that obtained from gel filtration and preliminary mass spectroscopy analysis. These are several possibilities to account for this difference. One possibility is that *Cry j* II protein is processed. It is probable that the N-terminal and C-terminal of the protein are cleaved. It is not clear at the present time whether this processing occurs in the cell or due to proteolysis during purification even though four different protease inhibitors were added in most of the purification steps. Nevertheless, the two N-terminal sequences obtained from the purified *Cry j* II (Example 2) also contained the N-terminal sequence (10 amino acid) published by Sakaguchi et al (*Allergy*, 45:309-312(1990)) suggesting that the N-terminal of *Cry j* II is probably hydrolyzed. Since Sakaguchi et al. (supra), did not use any protease inhibitors in their purification, a higher degree of hydrolysis might have occurred. This could explain why the N-terminal amino acid sequence that Sakaguchi et al. obtained was downstream of the N-terminal sequences as discussed in Example 2.

Another approach which may be used to purify native *Cry j* II or recombinant *Cry j* II is immunoaffinity chromatography. This technique provides a very selective protein purification due to the specificity of the interaction between monoclonal antibodies and antigen. Murine polyclonal and monoclonal antibodies are generated against purified *Cry j* II. These antibodies are used for purification, characterization, analysis and diagnosis of the allergen *Cry j* II.

Example 2

Protein Sequencing of Purified *Cry j* II

Cry j II protein was isolated as in Example 1. The doublet band shown on

SDS-PAGE (Fig. 1a) was electroblotted onto ProBlott (Applied Biosystems, Foster City, CA). Sequencing was performed with the Beckman/Porton Microsequencer (model LF3000, Beckman Instruments, Carlsbad, CA), a Programmable Solvent Module (Beckman System Gold Model 126, Beckman Instuments, Carlsbad, CA) and a Diode Array Detector Module for PTH-amino acid detection (Beckman System Gold Model 168, Beckman Instruments, Carlsbad, CA) following manufacturers specifications.

A single N-terminal sequence analysis of the upper doublet band and multiple N-terminal sequence analyses of the lower doublet band showed that both bands contained two N-termini, designated "long" and "short". The lower doublet band contained approximately 3.3 picomoles of the long form and 8.3 picomoles of the short form. This difference in yields was sufficient to make sequence assignments according to the quantitation at each sequencer cycle. The upper doublet band contained approximately 8.3 picomoles of both sequences. The revealed long sequence was NH_2-RKVEHSRHDAINIFNVEKYGAVGDGKH-DCTEAFSTAW(Q) ( ) ( ) ( ) KNP ( ) -COOH, (SEQ ID NO: 4) where (Q) indicates a tentative identification of glutamine at position 38 and () indicated unknown residues at positions 39-41 and 45. The revealed "short" sequence was NH_2-SRHDAINIFNVEKYGAVGDGKHDCTEAFSTAWS-COOH (SEQ ID NO: 5). Thus the long *Cry j* II sequence had five additional amino terminal residues than the short form and the sequence of the short form exactly matched that of the long form. In addition, both the long and short forms of *Cry j* II contained the ten amino acids, NH_2-AINIFNVEKY-COOH (SEQ ID NO: 6), previously described for *Cry j* II (Sakaguchi et al. 1990, supra). The previously published ten amino acids (Sakaguchi et al. 1990, supra) correspond to amino acids ten through 19 of the long form described above.

Example 3

Extraction of RNA From Japanese Cedar Pollen and Staminate Cones and Cloning of *Cry j* II

Fresh pollen and staminate cone samples, collected from a single *Cryptomeria japonica* (Japanese Cedar) tree at the Arnold Arboretum (Boston, MA), were frozen immediately on dry ice. RNA was prepared from 500 mg of each sample, essentially as described by Frankis and Mascarhenas (1980) Ann. Bot. 45: 595-599. The samples were ground by mortar and pestle on dry ice and suspended in 5 ml of 50 mM Tris pH 9.0 with 0.2 M NaCl, 1 mM EDTA, 0.1% SDS that had been treated overnight with 0.1% diethyl pyrocarbonate (DEPC). After five

extractions with phenol/chloroform/isoamyl alcohol (mixed 25:24:1), the RNA was precipitated from the aqueous phase with 0.1 volume 3M sodium acetate and 2 volumes ethanol. The pellets were recovered by centrifugation, resuspended in 2 ml dH_2O and heated to 65°C for 5 minutes. Two ml 4M lithium chloride was added to the preparation and the RNA was precipitated overnight at 0°C. The RNA pellets were recovered by centrifugation, resuspended in 1 ml dH_2O, and again precipitated with 3M sodium acetate and ethanol on dry ice for one hour. The final pellet was washed with 70% ethanol, air dried and resuspended in 100 µl DEPC-treated dH_2O and stored at -80°C.

Double stranded cDNA was synthesized from 4 µg pollen RNA or 8 µg flowerhead RNA using a commercially available kit (cDNA Synthesis System kit, BRL, Gaithersburg, MD). The double-stranded cDNA was phenol extracted, ethanol precipitated, blunted with T4 DNA polymerase (Promega, Madison, WI), and then ligated to ethanol precipitated, self annealed, AT and AL oligonucleotides for use in a modified Anchored PCR reaction, according to the method of Rafnar *et al.* (1990) J. Biol. Chem. 266: 1229-1236 ; Frohman *et al.* (1990) Proc. Natl. Acad. Sci. USA 85: 8998-9002; and Roux *et al.* (1990) BioTech. 8: 48-57. Oligonucleotide AT has the sequence (SEQ ID NO: 10) 5'-GGGTCTAGAGGTACCG-TCCGTCCGATCGATCATT-3' (Rafnar *et al.* supra). Oligonucleotide AL has the sequence (SEQ ID NO: 11) 5'-AATGATCGATGCT (Rafnar *et al.* supra).

The first attempts at amplifying the amino terminus of *Cry j* II from the linkered cDNA (2 µl of a 20 µl reaction) was made using the degenerate oligonucleotide CP-11 and oligonucleotide AP. CP-11 has the sequence (SEQ ID NO: 12) 5'-ATACTTCTCIACGTTGAA-3', wherein A at positon 1 can be G, C at position 4 can be T, C at position 7 can be T, I at position 10 is inosine to reduce degeneracy (Knoth et al. (1988) *Nucleic Acids Res.* 16: 10932), G at position 13 can be A, and G at position 16 can be A). AP, which has the sequence (SEQ ID NO: 13) 5'-GGGTCTAGAGGTA-CCGTCCG-3', corresponds to nucleotides 1 through 20 of the oligonucleotide AT. CP-11 is the degenerate oligonucleotide sequence that is complementary to the coding strand sequence substantially encoding amino acids PheAsnValGluLysTyr (SEQ ID NO: 14) (amino acids 59 to 64 of Fig. 4), which correspond to the carboxy terminus of the previously published *Cry j* II sequence (Sakaguchi et al., supra) shown in Fig. 4. All oligonucleotides were synthesized by Research Genetics Inc., Huntsville, AL.

Polymerase chain reactions (PCR) were carried out using a commercially available kit (GeneAmp DNA Amplification kit, Perkin Elmer Cetus, Norwalk, CT)

whereby 10 μl 10x buffer containing dNTPs was mixed with 100 pmoles of each oligonucleotide, cDNA (3-5 μl of a 20 μl first strand cDNA reaction mix), 0.5 μl Amplitaq DNA polymerase, and distilled water to 100 μl.

The samples were amplified with a programmable thermal controller (MJ Research, Inc., Cambridge, MA). The first 5 rounds of amplification consisted of denaturation at 94°C for 1 min, annealing of primers to the template at 45°C for 1 min, and chain elongation at 72°C for 1 min. The final 20 rounds of amplification consisted of denaturation as above, annealing at 55°C for 1 min, and elongation as above. The primary PCR reaction was carried out with 100 pmol each of the oligonucleotides AP and CP-11. Five percent (5 μl) of this initial amplification was then used in a secondary amplification with 100 pmoles each of AP and CP-12. CP-12 has the sequence (SEQ ID NO: 15) 5'-CCTGCAGTACTTCT-CIACGTTGAAIAT-3', wherein C at position 10 can be T, C at position 13 can be T, I at positions 16 and 25 are inosines to reduce degeneracy as above, G at position 19 can be A, and G at position 22 can be A. The sequence (SEQ ID NO: 16) 5'-CCTGCAG-3' (bases 1 through 7 of CP-12) represents a *Pst* I site added for cloning purposes; the remaining degenerate oligonucleotide sequence is complementary to the coding strand sequence that substantially encodes the amino acids IlePheAsnValGluLysTyr (SEQ ID NO: 17) (amino acids 58-64 of Fig. 4). Amplified DNA was recovered by sequential chloroform, phenol, and chloroform extractions, followed by precipitation on dry ice with 0.5 volumes of 7.5M ammonium acetate and 1.5 volumes of isopropanol. After precipitation and washing with 70% ethanol, the DNA was simultaneously digested with *Xba* I and *Pst* I in a 50 μl reaction, precipitated to reduce the volume to 10 μl, and electrophoresed through a preparative 2% GTG NuSeive low melt gel (FMC, Rockport, ME). The appropriate sized DNA area was visualized by ethidium bromide (EtBr) staining, excised, and ligated into appropriately digested pUC19 for sequencing by the dideoxy chain termination method of Sanger *et al.* (1977) Proc. Natl. Acad. Sci. USA 74: 5463-5476) using a commercially available sequencing kit (Sequenase kit, U.S. Biochemicals, Cleveland, OH). All resultant clones were sequenced, and none were found to contain *Cry j* II sequence. An alternate 2° PCR reaction was performed with AP and the nested oligonucleotide CP-21. CP-21 has the sequence (SEQ ID NO: 18) 5'-CCTGCAGTACTTCTCIACGTTGAAGAT-3' wherein C at position 10 can be T, C at position 13 can be T, I at position 16 is inosine to reduce degeneracy as above, G at position 19 can be A, G at position 22 can be A, and G at position 25 can be A or T. The sequence (SEQ ID NO: 16) 5'-CCTGCAG-3' (bases 1 through 7 of CP-21) represent a *Pst* I site added for cloning purposes; the

remaining degenerate oligonucleotide sequence is the non-coding strand sequence corresponding to coding strand sequence substantially encoding amino acids IlePheAsnValGluLysTyr (SEQ ID NO: 17) (amino acids 58 to 64 of Fig. 4).

A primary PCR was also performed on double-stranded, linkered cDNA using CP-23D and AP, as above, to attempt to amplify the 3' end of the *Cry j* II cDNA. A secondary PCR was performed using 5% of the primary reaction, using CP-24D and AP. CP-23D (sequence (SEQ ID NO: 19) 5'-GCIATTAATATTTTTAA-3', wherein the T at position 6 can be C or A, T at position 9 can be C, T at position 12 can be C or A, and T at position 15 can be C) is the coding strand sequence substantially encoding amino acids AlaIleAsnIlePheAsn (SEQ ID NO: 20) (amino acids 55 to 60 of Fig. 4); CP-24D (SEQ ID NO: 21) (sequence 5'-GGAATTCCGCIATTAATATTTTTAATGT-3', wherein the T at position 14 can be C or A, T at position 17 can be C, T at position 20 can be C or A, T at position 23 can be C, and T at position 26 can be C) contains the sequence 5'-GGAATTCC-3' (SEQ ID NO: 22) (bases 1 through 8 of CP-24), which represents an *Eco RI* site added for cloning purposes. The remaining degenerate oligonucleotide sequence of CP-24D substantially encodes amino acids AlaIleAsnIlePheAsnVal (SEQ ID NO: 23) (amino acids 55 to 61 of Fig. 4). Again, multiple clones were sequenced, none of which could be identified as *Cry j* II, and this approach was not pursued further.

Upon the characterization of novel *Cry j* II protein sequence data described in Example 2, new degenerate oligonucleotides for cloning *Cry j* II were designed and synthesized. All oligonucleotides mentioned hereafter were synthesized on an ABI 394 DNA/RNA Synthesizer (Applied Biosystems, Foster City, CA), and purified on NAP-10 columns (Pharmacia, Uppsala, Sweden) as per the manufacturers' instructions. Degenerate oligonucleotide CP-35 was used with AP on the double-stranded linkered cDNA in a primary PCR reaction carried out as described herein. CP-35 has the sequence (SEQ ID NO: 24) 5'-GCTTCGGTACAATCATGTTT-3', wherein T at position 3 can also be C; G at position 6 can also be A, T or C; A at position 9 can also be G; A at position 12 can also be G; A at position 15 can be G; and T at position 18 can also be C; this degenerate oligonucleotide sequence is the non-coding strand sequence corresponding to coding strand sequence substantially encoding amino acids LysHisAspCysThrGluAla of *Cry j* II (SEQ ID NO: 25) (amino acids 71 to 77 of Fig. 4). Five percent (5 μl) of this initial amplification, designated JC136, was then used in a secondary amplification with 100 pmoles each of AP and degenerate *Cry j* II primer CP-36, an internally nested *Cry j* II oligonucleotide primer with the sequence (SEQ ID NO: 26) 5'-

GGCTGCAGGTACAATCATGTTTGCCATC-3' wherein A at position 11 can also be G; A at position 14 can also be G; A at position 17 can also be G; T at position 20 can also be C; G at position 23 can also be A, T, or C; and A at position 26 can also be G. The nucleotides 5'-GGCTGCAG-3' (SEQ ID NO: 27) (bases 1 through 8 of CP-36) represent a *Pst* I restriction site added for cloning purposes. The remaining degenerate oligonucleotide sequence of CP-36 is the non-coding strand sequence corresponding to coding strand sequence substantially encoding amino acids AspGlyLysHisAspCysThr of *Cry j* II (SEQ ID NO: 28) (amino acids 69 to 75 of Fig. 4). The dominant amplified product, designated JC137, was a DNA band of approximately 265 base pairs, as visualized on an EtBr-stained 2% GTG agarose gel.

Amplified DNA was recovered by sequential chloroform, phenol, and chloroform extractions, followed by precipitation at -20°C with 0.5 volumes of 7.5 ammonium acetate and 1.5 volumes of isopropanol. After precipitation and washing with 70% ethanol, the DNA was simultaneously digested with *Xba* I and *Pst* I in a 15 μl reaction and electrophoresed through a preparative 2% GTG SeaPlaque low melt gel (FMC, Rockport, ME). The appropriate sized DNA band was visualized by EtBr staining, excised, and ligated into appropriately digested pUC19 for sequencing by the dideoxy chain termination method (Sanger et al. (1977) *Proc. Natl Acad Sci. USA* 74: 5463-5476) using a commercially available sequencing kit (Sequenase kit, U.S. Biochemicals, Cleveland, OH).

The clones designated pUC19JC137a, pUC19JC137b, and pUC19JC137e were found to contain sequences encoding the amino terminus of *Cry j* II. All three clones had identical sequence in their regions of overlap, although all three clones had different lengths in the 5' untranslated region. Clone pUC19JC137b was the longest clone. The translated sequence of these clones had complete identity to the disclosed 10 amino acid sequence of *Cry j* II (Sakaguchi et al., supra.), as well as to the *Cry j* II amino acid sequence described in Example 2. Amino acid numbering is based on the sequence of the full length protein; amino acid 1 corresponds to the initiating methionine (Met) of *Cry j* II. The position of the initiating Met was supported by the presence of an upstream in-frame-stop codon and by 78% homology of the surrounding nucleotide sequence with the plant consensus sequence that encompasses the initiating Met, as reported by Lutcke et al. (1987) *EMBO J.* 6:43-48.

The cDNA encoding the remainder of *Cry j* II gene was cloned from the linkered cDNA by using oligonucleotides CP-37 (SEQ ID NO: 29) (which has the sequence 5'-ATGTTGGACAGTGTTGTCGAA-3') and AP in a primary PCR, designated JC138ii. Oligonucleotide CP-37 corresponds to nucleotides 129 to 149 of

Fig. 4, and is based on the nucleotide sequence determined for the partial *Cry j* II clone pUC19JC137b.

A secondary PCR reaction was performed on 5% of the initial amplification mixture, with 100 pmoles each of AP and CP-38 (SEQ ID NO: 30) (which has the sequence 5'-GGGAATTCAGAAAAGTTGAGCATTCTCGT-3'), the nested primer. The nucleotide sequence (SEQ ID NO: 31) 5'-GGGAATTC-3' (bases 1 through 8 of CP-38) represents an *Eco* RI restriction site added for cloning purposes. The remaining oligonucleotide sequence corresponds to nucleotides 177 to 197 of Fig. 4, and is based on the nucleotide sequence determined for the partial *Cry j* II clone pUC19JC137b. The amplified DNA product, designated JC140iii, was purified and precipitated as above, followed by digestion with *Eco* RI and *Asp* 718 and electrophoresis through a preparative 1% low melt gel. The dominant DNA band, which was approximately 1.55 kb in length, was excised and ligated into pUC19 for sequencing. DNA was sequenced by the dideoxy chain termination method (Sanger et al. supra) using a commercially available kit (sequenase kit (U.S. Biochemicals, Cleveland, OH). Both strands were completely sequenced using M13 forward and reverse primers (N.E. Biolabs, Beverly, MA) and internal sequencing primers CP-35, CP-38, CP-40, CP-41, CP-42, CP-43, CP-44, CP-45, CP-46, CP-47, CP-48, CP-49,CP-50, and CP-51. CP-40 (SEQ ID NO: 32) has the sequence 5'-GTTCTTCAATGGGCCATGT-3' and corresponds to nucleotides 359 to 377 of Fig. 4. CP-41 (SEQ ID NO: 33) has the sequence 5'- GTGTTAGGACT-GTCTCTCGG-3', which is the non-coding strand sequence that corresponds to nucleotides 720 to 739 of Fig. 4. CP-42 (SEQ ID NO: 35) has the sequence 5'-TGTCCAGGCCAT-GGAATAAG-3', which corresponds to nucleotides 864 to 883 of Fig. 4 except that the first nucleotide was synthesized as a T rather than the correct G. CP-43 has the sequence (SEQ ID NO: 35) 5'-GCCTTACATGGACTGCAACC-3', which is the non-coding strand sequence that corresponds to nucleotides 1476 to 1495 of Fig. 4. CP-44 has the sequence (SEQ ID NO: 36) 5'-TCCACGGGTCTGATAATCCA-3', which corresponds to nucleotides 612 to 631 of Fig. 4. CP-45 has the sequence (SEQ ID NO: 37) 5'-AGGCAGGAAGCAATTTT-CCC-3', which is the non-coding strand sequence that corresponds to nucleotides 1254 to 1273 of Fig. 4. CP-46 has the sequence (SEQ ID NO: 38) 5'-TACTGCACTTCAGCT-TCTGC-3', which corresponds to nucleotides 1077 to 1096 of Fig. 4. CP-47 has the sequence (SEQ ID NO: 39) 5'-GGGGGTCTCCGAATTTATCA-3', which is the non-coding strand sequence that substantially corresponds to nucleotides 1039 to 1058 of Fig. 4, except that the fifth nucleotide of CP-47 was synthesized as a G rather than the correct nucleotide, T.

CP-48 (SEQ ID NO: 40), which has the sequence 5'-GGATATTTCAGTGGACACGT-3', corresponds to nucleotides 1290 to 1309 of Fig. 4. CP-49 (SEQ ID NO: 41) has the sequence 5'-TATTAGAAGACC-CTGTGCCT-3', which is the non-coding strand sequence that corresponds to nucleotides 821 to 840 of Fig. 4. CP-50 (SEQ ID NO: 42) has the sequence 5'-CCATGTAAGGCCAAGTTAGT-3', which corresponds to nucleotides 1485 to 1504 of Fig. 4. CP-51 (SEQ ID NO: 43) has the sequence 5'-ACACCTTTACCCATTAGAGT-3', which is the non-coding strand sequence that corresponds to nucleotides 486 to 505 of Fig. 4.

Three clones, designated pUC19JC140iiia, pUC19JC140iiid and pUC19JC140iiie, were subsequently found to contain partial *Cry j* II sequence. The sequence of clone pUC19JC140iiid was chosen as the consensus sequence since it had the longest 3' untranslated region. The sequences of pUC19JC140iiid and pUC19JC137b were used to construct the composite *Cry j* II sequence shown in Fig. 4. In this composite, nucleotide 230 is reported as the A found in pUC19JC137b (also, pUC19JC137a, pUC19JC140iiia and pUC19JC140iiie) not as the G found in pUC19JC140iiid; however both A and G at nucleotide 230 encode Lys at amino acid 63. The sequence of clone pUC19JC140iiia was identical to that of pUC19JC140iiid except for the following: pUC19JC140iiia has a T at nucleotide 357 in place of a C (no predicted change in amino acid 106), has C at nucleotide 754 instead of T (changes amino acid 238 from Ile to Thr), C at nucleotide 1246 instead of T (changes amino acid 402 from Leu to Pro), and T at nucleotide 1672 instead of C (untranslated region). The sequence of clone pUC19JC140iiie was identical to that of pUC19JC140iiid except for G at nucleotide 794 instead of A (changes amino acid 251 from Ile to Met), and T at nucleotide 357 in place of C (no predicted change in amino acid 106).

An earlier attempt at cloning the JC140iii PCR product using an *Eco* RI/*Xba* I digest (oligonucleotide AP has both *Xba* I and *Asp* 718 restriction enzyme sites) yielded cDNA that was cut in half due to an internal Xba I restriction site in the *Cry j* II cDNA, giving rise to 800 and 750 bp bands; the 750 bp band was succesfully cloned into *Eco* RI/*Xba* I digested pUC19 and sequenced. Two 750 bp clones were sequenced and found to be the 5' half of the *Cry j* II molecule: clones pUC19JC140-2a and pUC19JC140-2b. Clone pUC19JC140-2a has C for nucloeotide 297 instead of T (changes amino acid 86 from Cys to Arg) and clone pUC19JC140-2b has G for nucleotide 753 instead of A (changes amino acid 238 from Ile to Val). Both clone pUC19JC140-2a and clone pUC19JC140-2b have a T at nucleotide 357 in place of C (no predicted change in amino acid 106).

Two different PCR amplifications were also sequenced directly to verify the clonal *Cry j* II sequence using the Amplitaq Cycle Sequencing kit (Perkin Elmer Cetus, Norwalk, CT). This procedure involves the [^{32}P]-end-labelling of oligonucleotide sequencing primers which are then annealled (1.6 pmoles in 1 μl) to template DNA and elongated with dideoxy NTPs (methodology of Sanger et al. (1977) *Proc. Natl. Acad. Sci. USA* 74:5463-5476) in a PCR reaction also containing 4 μl 10X Cycling Mix (contains 0.5 U/μl Amplitaq DNA Polymerase), 5 μl template DNA (10-100 fmoles) and dH_2O to 20 μl . The dGTP in the termination mixes in this kit have been replaced by 7-deaza-dGTP, which provides increased resolution of sequences containing high G+C regions of DNA. The template DNA was a PCR product that was recovered by sequential chloroform, phenol, and chloroform extractions, precipitated at -20°C with 0.5 volumes of 7.5 ammonium acetate and 1.5 volumes of isopropanol, then electrophoresed through a preparative 1 or 2% SeaPlaque low melt gel (FMC). Appropriate sized DNA bands were visualized by EtBr staining, excised, and treated with Gelase (Epicentre Technologies, Madison, WI) to remove the agarose. The DNA was again precipitated, and resuspended in 50 μl TE (10 mM Tris, pH 7.4, 1 mM EDTA, pH 8.0) containing 20 μg/ml RNAse (Boehringer Mannheim, Indianapolis, IN). Two secondary amplifications which had been used to clone *Cry j* II were repeated, and used as template DNA for PCR cycle sequencing: JC137ii, the 5' end PCR, (amplified from the 1° PCR JC136 above) was reamplified with oligonucleotides AP and CP-36; and JC140ii, the 3' end PCR, (amplified from the 1° PCR JC138ii above) was reamplified with oligonucleotides AP and CP-38. Both of the 1° amplifications used were precipitated, electrophoresed through a preparative 1 or 2% SeaPlaque low melt gel (FMC), and the appropriate sized bands were visualized by EtBr staining and excised. Two μl of each 1° amplification was then used in the corresponding 2° PCR reaction. The 2° PCR product was then prepared as DNA template for PCR cycle sequencing as described above. The oligonucleotides used as primers in PCR cycle sequencing, many of which were used to sequence the clones, are as follows: for JC137ii, CP-36 and CP-39 (SEQ ID NO: 44), which has the sequence 5'-CTGTCCAACATAATTTGGGC-3' and is the non-coding strand sequence corresponding to nucleotides 120 to 139 of Fig. 4. The oligonucleotide primers used for sequencing JC140ii were CP-38, CP-40, CP-41, CP-42, CP-43, CP-44, CP-45, CP-46, CP-47, CP-49, CP-50, CP-54 (SEQ ID NO: 45), which has the sequence 5'-CATGGCAGGGTGGTTCAGGC-3', corresponds to nucleotides 985 to 1004 of Fig. 4, CP-55 (SEQ ID NO: 46), which has the sequence 5'-TAGCCCCATTTACGTGCACG-3' and is the non-coding strand sequence that

corresponds to nucleotides 929 to 948 of Fig. 4, and CP-56 (SEQ ID NO: 47), which has the sequence 5'-TTGGGGTCGAGGCCTCCGAA-3' and corresponds to nucleotides 1437 to 1456 of Fig. 4. The sequence of this full-length PCR cycle sequencing had only 2 nucleotide changes from the composite pUC19JC137b/pUC19JC140iiid *Cry j* II sequence shown in Figure 4, neither of which lead to an amino acid change. There was a T instead of C at nucleotide 357 (no predicted change in amino acid 106), and a C instead of A at nucleotide 635 (no amino acid change).

The nucleotide and predicted amino acid sequences of *Cry j* II are shown in Figs. 4 and 5. This is a composite nucleotide sequence from the two overlapping clones pUC19JC137b and pUC19JC140iiid. Sequencing of multiple independent clones and cycle sequencing of PCR product confirmed the nucleotide sequence of Figure 4. There were several nucleotide changes resulting in predicted amino acid changes, as cited above. However, all nucleotide polymorphisms, with the exception of the T for C substitition at nucleotide 357, were only observed in single clones or sequencing reactions. Although T was seen at nucleotide 357 in all clones except pUC19JC140iiid, both C and T encode Leu at amino acid 106.

The complete cDNA sequence for *Cry j* II is composed of 1726 nucleotides, including 41 nucleotides of 5' untranslated sequence, an open reading frame of 1542 nucleotides starting with the codon for an initiating Met (nucleotides 42-44 of Fig. 4), and a 143 bp 3' untranslated region. There is a consensus polyadenylation signal sequence in the 3' untranslated region 64 nucleotides 5' to the poly A tail (nucleotides 1654-1659 of Fig. 4). The position of the initiating Met is confirmed by the presence of an in-frame upstream stop codon and by 78% homology with the plant consensus sequence that encompasses the initiating Met (TAAAAUGGC (bases 38 through 46 of Fig. 4 (SEQ ID NO: 48)) found in *Cry j* II compared with the AACAAUGGC (SEQ ID NO: 49) consensus sequence for plants, Lutcke et al. (1987) *EMBO J.* 6: 43-48). The open reading frame encodes a deduced protein of 514 amino acids that has complete sequence identity with the published partial protein sequence for *Cry j* II (Sakaguchi et al. supra), which corresponds to amino acids 55 through 64 of Fig. 4. The predicted *Cry j* II protein has 20 Cys, contains four potential N-linked glycosylation sites corresponding to the consensus sequence N-X-S/T, has a predicted molecular weight of 56.6 kDa and a predicted pI of 9.08.

Detection of three separate NH_2 termini sequences for *Cry j* II (the long form and the short form as determined in Example 2 and the NH_2 terminus determined by Sakaguchi et al., supra, as shown in Fig. 6) may suggest that the amino terminus of the mature *Cry j* II protein is blocked and that the sequences obtained by sequence

analysis of purified protein represent proteolytic cleavage products. As shown in Fig. 6, the amino acid sequence of the long form of *Cry j* II begins at amino acid 46 and the amino acid sequence of the short form of *Cry j* II begins at amino acid 51; and the NH2-terminal sequence determed by Sakaguchi et al. begins at amino acid 54. It is also possible that amino acids 1 to 45 represent the leader/pre-pro position of *Cry j* II that is enzymatically cleaved to give a functionally active protein beginning at amino acid 46 of Fig. 4. The sequences beginning at amino acids 51 and 54 represent breakdown products of the protein beginning at amino acid 46. There is a predicted cleavage site between amino acids 22 and 23 of Fig. 4 using the method of von Heijne (Nucleic Acids Res. (1986) 14:4683-4690). If the mature *Cry j* II protein started at amino acid 23 in Fig. 4, the protein would be 492 amino acids long with a predicted molecular weight of 54.2 kDa and a predicted pI of 9.0.

Searching the Swiss-Prot data base with the *Cry j* II sequence demonstrated that *Cry j* II is 43.3% homologous (33.3% identical to polygalacturonase of tomato (*Lycopersicon esculentum*) and 48.4% homologous (32.6% identical) to polygalacturonase of corn, Zea mays. All nucleotide and amino acid sequence analyses were performed using PCGENE (Intelligenetics, Mountain View, CA.).

Example 4

Extracti n f RNA from Japanese Cedar Pollen Collected in Japan and Expression of Recombinant *Cry j* II

Fresh pollen collected from a pool of *Cryptomeria japonica* (Japanese cedar) trees in Japan was frozen immediately on dry ice. RNA was prepared from 500 mg of the pollen, essentially as described by Frankis and Mascarenhas *Ann. Bot.* 45:595-599. The samples were ground by mortar and pestle on dry ice and suspended in 5 ml of 50 mM Tris pH 9.0 with 0.2 M NaCl, 1 mM EDTA, 1% SDS that had been treated overnight with 0.1% DEPC. After five extractions with phenol/chloroform /isoamyl alcohol (mixed at 25:24:1), the RNA was precipitated from the aqueous phase with 0.1 volume 3 M sodium acetate and 2 volumes ethanol. The pellets were recovered by centrifugation, resuspended in 2 ml dH_20 and heated to 65°C for 5 minutes. Two ml of 4 M lithium chloride were added to the RNA preparations and they were incubated overnight at 0°C. The RNA pellets were recovered by centrifugation, resuspended in 1 ml dH_20, and again precipitated with 3 M sodium acetate and ethanol overnight. The final pellets were resuspended in 100 µl dH_20 and stored at -80°C.

Double stranded cDNA was synthesized from 8 µg pollen RNA using the cDNA Synthesis Systems kit (BRL) with oligo dT priming according to the method of Gubler and Hoffman (1983) *Gene* 25:263-269. PCRs were carried out using the GeneAmp DNA Amplification kit (Perkin Elmer Cetus) whereby 10 µl 10x buffer containing dNTPs was mixed with 100 pmol each of a sense oligonucleotide and an anti-sense oligonucleotide, cDNA (10 µl of a 400 µl double stranded cDNA reaction mix), 0.5 µl Amplitaq DNA polymerase, and distilled water to 100 µl.

The samples were amplified with a programmable thermal controller from MJ Research, Inc. (Cambridge, MA). The first 5 rounds of amplification consisted of denaturation at 94°C for 1 min, annealing of primers to the template at 45°C for 1 min, and chain elongation at 72°C for 1 min. The final 20 rounds of amplification consisted of denaturation as above, annealing at 55°C for 1 min, and elongation as above.

A new set of primer pairs was synthesized for amplification of a *Cry j* II cDNA from the initiating Met to the stop codon. CP-52 (SEQ ID NO: 50) has the sequence 5'- GCCGAATTCATGGCCATGAAATTAATT-3' where the nucleotide sequence 5'-GCCGAATTC-3' (SEQ ID NO: 51) (bases 1 through 9 of CP-52 represents an *Eco* RI restriction site added for cloning purposes, and the remaining sequence corresponds to nucleotides 42 to 59 of Fig. 4. CP-53 (SEQ ID NO: 52)

has the sequence 5'-CGGGGATCCTCATTATGGATG-GTAGAT-3' where the nucleotide sequence 5'-CGGGGATCC-3' (SEQ ID NO: 53) (bases 1 through 9 of CP-53 represents a *Bam* HI restriction site added for cloning purposes, and the remaining oligonucleotide sequence of CP-53 is complementary to coding strand sequence corresponding to nucleotides 1572 to 1589 of Fig. 4. The PCR reaction with CP-52 and CP-53 on the double stranded Japanese Cedar pollen cDNA yielded a band of approximately 1.55 kb on an EtBr-stained agarose minigel, and was called JC145. Amplified DNA was recovered by sequential chloroform, phenol, and chloroform extractions, followed by precipitation at -20°C with 0.5 volumes of 7.5 ammonium acetate and 1.5 volumes of isopropanol. After precipitation and washing with 70% ethanol, the DNA was simultaneously digested with *Eco* RI and *Bam* HI in a 15 μl reaction, and electrophoresed through a preparative 1% SeaPlaque low melt gel (FMC). Appropriate sized DNA bands were visualized by EtBr staining, excised, and ligated into appropriately digested pUC19 for sequencing by the dideoxy chain termination method (Sanger et al. (1977) *Proc. Natl. Acad. Sci. USA* 74:5463-5476) using a commercially available sequencing kit (Sequenase kit, U.S. Biochemicals, Cleveland, OH).

Clones pUC19JC145a and pUC19JC145b were completely sequenced using M13 forward and reverse primers (N.E. Biolabs, Beverly, MA) and internal sequencing primers CP-41, CP-42, CP-44, CP-46, and CP-51. The nucleotide and deduced amino acid sequences of clones pUC19JC145a and pUC19JC145b were identical to the *Cry j* II sequence of Fig. 4, with the following exceptions. Clone pUC19JC145a was found to contain a single nucleotide difference from the previously known *Cry j* II sequence: it has a C at nucleotide position 1234 of Fig. 4 rather than the previously described T. This nucleotide change results in a predicted amino acid change from Ile to Thr at amino acid 398 of the *Cry j* II protein. Clone pUC19JC145b has a G at nucleotide position 1088 of Fig. 4 rather than the previously described A, and an A for a G at nucleotide 1339. The nucleotide change at 1088 is silent and does not result in a predicted amino acid change. The nucleotide change at position 1339 results in a predicted amino acid change from Ser to Asn at amino acid 433 of the *Cry j* II protein. None of these polymorphisms have yet been confirmed by independently-derived PCR clones or by direct amino acid sequencing and may be due to the inherent error rate of Taq polymerase (approximately 2 x 10^{-4}, Saiki et al. (1988) Science 239:487-491). However, such polymorphisms in primary nucleotide and amino acid sequences are expected.

Expression of *Cry j* II was performed as follows. Ten μg of pUC19JC145b was digested simultaneously with *Eco* RI and *Bam* HI. The nucleotide insert

encoding *Cry j* II (extending from nucleotide 42 through 1589 of Fig. 4) was isolated by electrophoresis of this digest through a 1% SeaPlaque low melt agarose gel. The insert was then ligated into the appropriately digested expression vector pET-11d (Novagen, Madison, WI; Jameel et al. (1990) *J. Virol.* 64:3963-3966) modified to contain a sequence encoding 6 histidines (His 6) immediately 3' of the ATG initiation codon followed by a unique *Eco* RI endonuclease restriction site. A second *Eco* RI endonuclease restriction site in the vector, along with neighboring *Cla* I and *Hind* III endonuclease restriction sites, had previously been removed by digestion with *Eco* RI and *Hind* III, blunting and religation. The histidine (His_6) sequence was added for affinity purification of the recombinant protein (*Cry j* I) on a Ni^{2+} chelating column (Hochuli et al. (1987) *J. Chromatog.* 411:177-184; Hochuli et al. (1988) *Bio/Tech.* 6:1321-1325.). A recombinant clone was used to transform *Escherichia coli* strain BL21-DE3, which harbors a plasmid that has an isopropyl-ß-D-thiogalactopyranoside (IPTG)-inducible promoter preceding the gene encoding T7 polymerase. Induction with IPTG leads to high levels of T7 polymerase expression, which is necessary for expression of the recombinant protein in pET-11d. Clone pET-11dΔHRhis$_6$JC145b.a was confirmed to be a *Cry j* II clone in the correct reading frame for expression by dideoxy sequencing (Sanger et al. supra) with CP-39.

Expression of the recombinant protein was examined in an initial small culture. An overnight culture of clone pET-11dΔHRhis$_6$JC145b.a was used to innoculate 50 ml of media (Brain Heart Infusion Media, Difco) containing ampicillin (200 µg/ml), grown to an A_{600} = 1.0 and then induced with IPTG (1 mM, final concentration) for 2 hrs. One ml aliquots of the bacteria were collected before and after induction, pelleted by centrifugation, and crude cell lysates prepared by boiling the pellets for 5 minutes in 50 mM Tris HCl, pH 6.8, 2 mM EDTA, 1% SDS, 1% ß-mercaptoethanol, 10% glycerol, 0.25% bromophenol blue (Studier et al., (1990) *Methods in Enzymology* 185:60-89). Recombinant protein expression was examined on a 12% Coomassie blue-stained SDS-PAGE gel, according to the method in Sambrook et al., supra, on which 25 µl of the crude lysates were loaded. A negative control consisted of crude lysate from uninduced bacteria containing the plasmid with *Cry j* II. There was no notable increase in production of any recombinant *E. coli* protein in the range of 58 Kd, the size predicted for the recombinant *Cry j* II with the His_6 leader.

The pET-11dΔHRhis$_6$JC145b.a clone was then grown on a larger scale to examine if there was any recombinant protein being expressed. A 2 ml culture of bacteria containing the recombinant plasmid was grown for 8 hr, then 3 µl was

spread onto each of 6 (100 x 15 mm) petri plates with 1.5% agarose in LB medium (Gibco-BRL, Gaithersburg, MD) containing 200 μg/ml ampicillin, grown to confluence overnight, then scraped into 6 L of liquid media (Brain Heart Infusion media, Difco) containing ampicillin (200 μg/ml). The culture was grown until the absorbance at A_{600} was 1.0, IPTG added (1 mM final concentration), and the culture grown for an additional 2 hours.

Bacteria were recovered by centrifugation (7,930 xg, 10 min) and lysed in 50 ml of 6M Guanidine-HCl, 0.1M Na_2HPO_4, pH 8.0, for 1 hour with vigorous shaking. Insoluble material was removed by centrifugation (11,000 xg, 10 min, 4° C). The pH of the lysate was adjusted to pH 8.0, and the lysate applied to a 50 ml Nickel NTA agarose column (Qiagen) that had been equilibrated with 6 M Guanidine HCl, 100 mM Na_2HPO_4, pH 8.0. The column was sequentially washed with 6 M Guanidine HCl, 100 mM Na_2HPO_4, 10 mM Tris-HCl, pH 8.0, then 8 M urea, 100 mM Na_2HPO_4, pH 8.0, and finally 8 M urea, 100 mM sodium acetate, 10 mM Tris-HCl, pH 6.3. The column was washed with each buffer until the flow through had an $A_{280} \leq 0.05$.

The recombinant *Cry j* II protein was eluted with 8 M urea, 100 mM sodium acetate, 10 mM Tris-HCl, pH 4.5, and collected in 10 ml aliquots. The protein concentration of each fraction was determined by A_{280} and the peak fractions pooled. An aliquot of the collected recombinant protein was analyzed on SDS-PAGE according to the method in Sambrook et al. supra.

This 6L prep, JCIIpET-1, yielded 1.5 mg of recombinant *Cry j* II, which was resolved into 2 major bands on SDS-PAGE at 58 kDa and 24 kDa. The 58 kDa band, which represents recombinant *Cry j* II, was approximately 9-10% of the total protein as determined by densitometry measurement (Shimadzu Flying Spot Scanner, Shimadzu Scientific Instruments, Inc., Braintree, MA). The 24 kDa band accounts for about 90% of the total protein and may represent a degradation product of the recombinant *Cry j* II or an *E. coli* contaminant.

Another *Cry j* II expression construct was made by the ligation of the pUC19JC140iiid *Cry j* II insert into appropriately digested pET11dΔHR (with the 6 histidine leader). The vector was derived from another pET11dΔHR construct whose insert supplied an EcoR I site (at the 5' pET11dΔHR-insert junction) and an Asp 718 site (at the 3' end of the insert); the construct was digested with these two enzymes, run on a low melt minigel as above, and the vector recovered as a band in low melt agarose. The pUC19JC140iiid construct was digested with Eco R I and Asp 718 to release the *Cry j* II insert, which was isolated on a low melt minigel and ligated into the Eco R I/Asp 718 digested pET11dΔHR vector prepared above. Five

clones were found to contain the correct nucleotide sequence at the insert/vector 5' junction, when sequenced by dideoxy sequencing (as above) with CP-39. This new construct, when expressed, would begin at amino acid 46 of *Cry j* II as shown in Figs. 4 and 5. This recombinant protein is designated r*Cry j* II Δ46. A 50 ml small scale expression test (as performed above) showed that the expression level of r*Cry j* II Δ46 from this construct, designated pET11dΔHRJC140iiid2, would be much greater than the initial expression level from pET11dΔHRJC145b2. A 9L prep, JCIIpET-3, was processed as above, and yielded 200 mg of r*Cry j* II Δ46 at 80% purity as determined by densitometry of a Coomasie blue stained 12% SDS-PAGE gel.

Example 5

Northern blot on RNA from Japanese Cedar Pollen Sources

A northern blot analysis was performed on the RNA isolated from Japanese Cedar pollen from both the Arnold Arboretum tree and the pooled trees from Japan. Using essentially the method of Sambrook, supra, ten μg of RNA isolated from Japanese cedar pollen collected from the Arnold Arboretum (Boston, MA) and 15 μg pooled RNA from Japanese cedar pollen collected from trees in Japan were run on a 1.2% agarose gel containing 38% formaldehyde and 1X MOPS (20X = 0.4M MOPS, 0.02M EDTA, 0.1M NaOAc, pH 7.0) solution. The RNA samples (first precipitated with 1/10 volume sodium acetate, 2 volumes ethanol to reduce volume and resuspended in 5.5 μl dH2O) were run with 10 μl formaldehyde/formamide buffer containing loading dyes with 15.5% formaldehyde, 42% formamide, and 1.3X MOPS solution, final concentration. The samples were transferred to Genescreen Plus (NEN Research Products, Boston, MA) by capillary transfer in 10X SSC (20X = 3M NaCl, 0.3M Sodium Citrate), after which the membrane was baked 2 hrs at 80°C and UV irradiated for 3 minutes. Prehybridization of the membrane was at 60°C for 1 hour in 4 ml 0.5M NaPo4 (pH 7.2), 1mM EDTA, 1% BSA, and 7% SDS. The antisense probe was synthesized by asymmetric PCR on the JC145 amplification in low melt agarose (above), where 2 μl DNA is amplified with 2 μl dNTP mix (0.167mM dATP, 0.167mM dTTP, 0.167mM dGTP, and 0.033mM dCTP), 2 μl 10X PCR buffer, 10 μl ^{32}P-dCTP (100 μCi; Amersham, Arlington Heights, Il), 1 μl (100 pmoles) antisense primer CP-53, 0.5 μl Taq polymerase, and dH2O to 20 μl; the 10X PCR buffer, dNTPs and Taq polymerase were from Perkin Elmer Cetus (Norwalk, CT). Amplification consisted of 30 rounds of denaturation at 94°C for 45 sec, annealing of primer to the template at 60°C for 45 sec, and chain elongation at 72 C for 1 min. The reaction was stopped by addition of 100 μl TE,

and the probe recovered over a 3cc G-50 spin column (2 ml G-50 Sephadex [Pharmacia, Uppsala, Sweden] in a 3cc syringe plugged with glass wool, equilibrated with TE) and counted on a 1500 TriCarb Liquid Scintillation Counter (Packard, Downers Grove, IL). The probe was added to the prehybridizing buffer at 10^6 cpm/ml and hybridization was carried out at 60°C for 16 hrs. The blot was washed in high stringency conditions: 3x15 min at 65°C with 0.2%SSC/1% SDS, followed by wrapping in plastic wrap and exposure to film at -80°C. A seven hour exposure of this Northern blot analysis revealed a single thick band at approximately 1.7 kb for both RNA collected from the Arboretum tree and the RNA collected from the pooled trees from Japan. This message is the expected size for *Cry j* II as predicted by PCR analysis of the cDNA.

Example 6

Direct binding assay of IgE to *Cry j* I, *Cry j* II and recombinant *Cry j* II.

Corning assay plates (#25882-96) were coated with *Cry j* I or *Cry j* II at 2 μg/mL or recombinant *Cry j* II preparation at 10 μg/mL (approximately 20% pure) in a volume of 50 μL overnight at 4°C. The coating antigens were removed and the wells were blocked with 0.5% gelatin, PVP (polyvinyl pyrolidine) 1 mg/ mL in PBS, 200 μL/well for 2 hours at room temperature. The anti-*Cry j* I monoclonal antibody, 4B11, was serially diluted in PBS-Tween 20 starting at a 1:1000 dilution. The human plasma were serially diluted in PBS-Tween at a starting dilution of 1:2. For this set 23 plasma samples from patients symptomatic for Japanese cedar pollen allergy chosen for IgE binding analysis. The first antibody incubation proceeded overnight at 4°C. Following three washes with PBS-Tween the second antibodies were added (goat anti-mouse Ig or goat anti-human IgE both at 1:2000) and incubated for two hours at room temperature at 100 μL/well. This solution was removed and streptavidin-HRPO diluted to 1:10,000, was added at 100μL/well. The color was allowed to develop for 2-5 minutes. The reaction was stopped by the addition of 100μL/well of 1M phosphoric acid. Plates were read on a Microplate IL310 Autoreader (Biotek Instruments, Winooski, VT) with a 450nm filter. The absorbance levels of duplicate wells were averaged. The graphed results (log of the dilution vs. absorbance) of the ELISA assays are shown in Figs. 7 to 15. The summary of the results are given in Fig. 16. A positive binding result, indicated by a plus sign is determined to be a reading of two-fold or greater above background (no first antibody) at the second dilution of plasma (1:6).

In Fig. 7 the binding response of the monoclonal antibody, 4B11, and seven patients' (Batch 1) plasma IgE is shown to purified *Cry j* I as the coating antigen.

The monoclonal antibody, raised against purified *Cry j* I shows a saturating level of binding for the whole dilution series. The individual patient samples show a variable response of IgE binding to the *Cry j* I preparation. One patient, #1034, has no detectable binding to this protein preparation. All the patient samples were obtained from individuals claiming to be symptomatic for Japanese cedar pollen allergy and the results of their MAST scores are shown in Fig. 16. Fig. 8 is a graph representing the binding of the same antibody set as in Fig. 7 to purified native *Cry j* II. The anti-*Cry j* I monoclonal antibody, 4B11, is negative on this preparation demonstrating lack of cross-reactivity between the two allergen antigens. In general, there is a lower overall response to this allergenic component of cedar pollen with more patient samples showing decreased binding. However, patient #1034, that was negative on *Cry j* I shows very strong reactivity to *Cry j* II. In the last antigen set, Fig. 9, using recombinant *Cry j* II (r*Cry j* II), monoclonal antibody 4B11 reactivity is negative and there is further reduction in binding of the human IgE samples compared to biochemically purified *Cry j* II. Two of the patients, #1143 and #1146, are clearly positive for IgE binding to the recombinant form of *Cry j* II although the patient that reacted the strongest to biochemically purified form is negative here, 1034. Figs. 10-15 represent the application of the same antigen sets for the direct binding analysis of the next sixteen patients designated patient Batch 2 and patient Batch 3 in Figs. 10-15.

The table shown in Fig. 16 summarizes both the MAST scores, performed in Japan on the plasma samples before shipment using a commercially available kit, and the direct ELISA results outlined above. Two patients were negative by the MAST assay, however, one of these patients, #1143, was positive on all the ELISA antigens. The number of positive responses for each antigen is shown and this represents a measure relative allergenicity of the different allergen preparations. These results demonstrate that *Cry j* II is an allergen as defined by human allergic patient IgE reactivity and that there are some patients who are not reactive to *Cry j* I but are reactive to *Cry j* II. The frequency of response in this population of patients is less to *Cry j* II than to *Cry j* I.

Example 7

Japanese Cedar Pollen Allergic Patient T Cell Studies with *Cry j* II and *Cry j* II Peptides.

Synthesis of *Cry j* II Peptides

Japanese cedar pollen *Cry j* II peptides designated Cry j IIA Cry j IIB were synthesized using standard Fmoc/tBoc synthetic chemistry and purified by Reverse Phase HPLC. The amino acid sequence of peptide Cry J IIA is FTFKVDGIIAAYQ (SEQ ID NO: 54) which corresponds to amino acids 116-128 as shown if Figs 4 and 5. The amino acid sequence of peptide Cry j IIB is NGYFSGHVIPACKN (SEQ ID NO: 55) which corresponds to amino acids 416-429 as shown in Figs 4 and 5. The peptide names are consistent throughout.

T Cell Responses to Japanese Cedar Pollen Antigen Peptides

Peripheral blood mononuclear cells (PBMC) were purified by lymphocyte separation medium (LSM) centrifugation of 60 ml of heparinized blood from one Japanese cedar pollen-allergic patient who exhibited clinical symptoms of seasonal rhinitis and was MAST and/or skin test positive for Japanese cedar pollen. Long term T cell lines were established by stimulation of 2 X 10^6 PBL/ml in bulk cultures of complete medium (RPMI-1640, 2 mM L-glutamine, 100 U/ml penicillin/streptomycin, $5x10^{-5}$M 2-mercaptoethanol, and 10 mM HEPES supplemented with 5% heat inactivated human AB serum) with 10 μg/ml of partially purified native *Cry j* II for 7 days at 37°C in a humidified 5% CO_2 incubator to select for *Cry j* II reactive T cells. This amount of priming antigen was determined to be optimal for the activation of T cells from most Japanese cedar pollen allergic patients. Viable cells were purified by LSM centrifugation and cultured in complete medium supplemented with 5 units recombinant human IL-2/ml and 5 units recombinant human IL-4/ml for up to three weeks until the cells no longer responded to lymphokines and were considered "rested". The ability of the T cells to proliferate to peptides Cry j IIA and Cry j IIB, recombinant *Cry j* II (r*Cry j* II), purified native *Cry j* II, or purified native *Cry j* I was then assessed. For assay, 2 X 10^4 rested cells were restimulated in the presence of 2 X 10^4 autologous Epstein-Barr virus (EBV)-transformed B cells (prepared as described below) (gamma-irradiated with 25,000 RADS) with 2-50 μg/ml of r*Cry j* II, purified native *Cry j* II, peptides Cry j IIA and Cry j IIB, of purified native *Cry j* I, in a volume of 200 μl complete medium in duplicate or triplicate wells in 96-well round bottom plates for 2-4 days. The optimal incubation was found to be 3 days. Each well then received 1 μCi tritiated thymidine for 16-20 hours. The counts incorporated were collected onto glass fiber filter mats and processed for liquid scintillation counting. The maximum response in a titration of each peptide is expressed as the stimulation index (S.I.). The S.I. is the counts per minute (CPM) incorporated by cells in response to peptide, divided by the CPM incorporated by cells in medium only. An S.I. value

equal to or greater than 2 times the background level is considered "positive" and indicates that the peptide contains a T cell epitope. The results of this assay indicated that peptides *Cr j* II, and *Cry j* IIB did noit appear to contain a T cell epitope for this particular allergenic patient. However, additional Japanese cedar pollen allergic patients will be tested in this assay system and one or both of these peptides may contain T cell epitopes for other allergic individuals.

Preparation f (EBV)-transformed B Cells for Use as Antigen Presenting Cells

Autologous EBV-transformed cell lines were γ-irradiated with 25,000 Rad and used as antigen presenting cells in secondary proliferation assays and secondary bulk stimulations. These cell lines were also used as a control in the immuno-fluorescence flow cytometry analysis. These EBV-transformed cell lines were made by incubating 5 X 10^6 PBL with 1 ml of B-59/8 Marmoset cell line (ATCC CRL1612, American Type Culture Collection, Rockville, MD) conditioned medium in the presence of 1 μg/ml phorbol 12-myristate 13-acetate (PMA) at 37°C for 60 minutes in 12 X 75 mm polypropylene round-bottom Falcon snap cap tubes (Becton Dickinson Labware, Lincoln Park, NJ). These cells were then diluted to 1.25 X 10^6 cells/ml in RPMI-1640 as described above except supplemented with 10% heat-inactivated fetal bovine serum and cultured in 200 μl aliquots in flat bottom culture plates until visible colonies were detected. They were then transferred to larger wells until the cell lines were established.

Although the invention has been described with reference to its preferred embodiments, other embodiments, can achieve the same results. Variations and modifications to the present invention will be obvious to those skilled in the art and it is intended to cover in the appended claims all such modification and equivalents and follow in the true spirit and scope of this invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: IMMULOGIC PHARMACEUTICAL CORPORATION
(B) STREET: 610 Lincoln Street
(C) CITY: Waltham
(D) STATE: MA
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 02154
(G) TELEPHONE: (617) 466-6000
(H) TELEFAX: (617)466-6040

(ii) TITLE OF INVENTION: Allergenic Proteins and Peptides From Japanese Cedar Pollen

(iii) NUMBER OF SEQUENCES: 55

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: ASCII (TEXT)

(v) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Vanstone, Darlene
(B) REGISTRATION NUMBER: 35,729
(C) REFERENCE/DOCKET NUMBER: IPC-033PC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617) 466-6000
(B) TELEFAX: (617) 466-6040

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1726 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 42..1586

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TGAGTTCGAG ACAAGTATAG AAAGAATTTT CTTTTATTAA A ATG GCC ATG AAA
53
                                                  Met Ala Met Lys
                                                    1

TTA ATT GCT CCA ATG GCC TTT CTG GCC ATG CAA TTG ATT ATA ATG GCG
101
Leu Ile Ala Pro Met Ala Phe Leu Ala Met Gln Leu Ile Ile Met Ala
  5                  10                  15                  20

GCA GCA GAA GAT CAA TCT GCC CAA ATT ATG TTG GAC AGT GTT GTC GAA
149
Ala Ala Glu Asp Gln Ser Ala Gln Ile Met Leu Asp Ser Val Val Glu
                  25                  30                  35

AAA TAT CTT AGA TCG AAT CGG AGT TTA AGA AAA GTT GAG CAT TCT CGT
197
Lys Tyr Leu Arg Ser Asn Arg Ser Leu Arg Lys Val Glu His Ser Arg
              40                  45                  50

CAT GAT GCT ATC AAC ATC TTC AAT GTG GAA AAG TAT GGC GCA GTA GGC
245
His Asp Ala Ile Asn Ile Phe Asn Val Glu Lys Tyr Gly Ala Val Gly
          55                  60                  65

GAT GGA AAG CAT GAT TGC ACT GAG GCA TTT TCA ACA GCA TGG CAA GCT
293
Asp Gly Lys His Asp Cys Thr Glu Ala Phe Ser Thr Ala Trp Gln Ala
      70                  75                  80

GCA TGC AAA AAC CCA TCA GCA ATG TTG CTT GTG CCA GGC AGC AAG AAA
341
Ala Cys Lys Asn Pro Ser Ala Met Leu Leu Val Pro Gly Ser Lys Lys
  85                  90                  95                  100

TTT GTT GTA AAC AAT CTG TTC TTC AAT GGG CCA TGT CAA CCT CAC TTT
389
Phe Val Val Asn Asn Leu Phe Phe Asn Gly Pro Cys Gln Pro His Phe
                  105                 110                 115

ACT TTT AAG GTA GAT GGG ATA ATA GCT GCG TAC CAA AAT CCA GCG AGC
437
Thr Phe Lys Val Asp Gly Ile Ile Ala Ala Tyr Gln Asn Pro Ala Ser
              120                 125                 130

TGG AAG AAT AAT AGA ATA TGG TTG CAG TTT GCT AAA CTT ACA GGT TTT
485
Trp Lys Asn Asn Arg Ile Trp Leu Gln Phe Ala Lys Leu Thr Gly Phe
          135                 140                 145

ACT CTA ATG GGT AAA GGT GTA ATT GAT GGG CAA GGA AAA CAA TGG TGG
533
Thr Leu Met Gly Lys Gly Val Ile Asp Gly Gln Gly Lys Gln Trp Trp
      150                 155                 160

GCT GGC CAA TGT AAA TGG GTC AAT GGA CGA GAA ATT TGC AAC GAT CGT
581
Ala Gly Gln Cys Lys Trp Val Asn Gly Arg Glu Ile Cys Asn Asp Arg
  165                 170                 175                 180

GAT AGA CCA ACA GCC ATT AAA TTC GAT TTT TCC ACG GGT CTG ATA ATC
```

629
Asp Arg Pro Thr Ala Ile Lys Phe Asp Phe Ser Thr Gly Leu Ile Ile
185 190 195

CAA GGA CTG AAA CTA ATG AAC AGT CCC GAA TTT CAT TTA GTT TTT GGG
677
Gln Gly Leu Lys Leu Met Asn Ser Pro Glu Phe His Leu Val Phe Gly
200 205 210

AAT TGT GAG GGA GTA AAA ATC ATC GGC ATT AGT ATT ACG GCA CCG AGA
725
Asn Cys Glu Gly Val Lys Ile Ile Gly Ile Ser Ile Thr Ala Pro Arg
215 220 225

GAC AGT CCT AAC ACT GAT GGA ATT GAT ATC TTT GCA TCT AAA AAC TTT
773
Asp Ser Pro Asn Thr Asp Gly Ile Asp Ile Phe Ala Ser Lys Asn Phe
230 235 240

CAC TTA CAA AAG AAC ACG ATA GGA ACA GGG GAT GAC TGC GTC GCT ATA
821
His Leu Gln Lys Asn Thr Ile Gly Thr Gly Asp Asp Cys Val Ala Ile
245 250 255 260

GGC ACA GGG TCT TCT AAT ATT GTG ATT GAG GAT CTG ATT TGC GGT CCA
869
Gly Thr Gly Ser Ser Asn Ile Val Ile Glu Asp Leu Ile Cys Gly Pro
265 270 275

GGC CAT GGA ATA AGT ATA GGA AGT CTT GGG AGG GAA AAC TCT AGA GCA
917
Gly His Gly Ile Ser Ile Gly Ser Leu Gly Arg Glu Asn Ser Arg Ala
280 285 290

GAG GTT TCA TAC GTG CAC GTA AAT GGG GCT AAA TTC ATA GAC ACA CAA
965
Glu Val Ser Tyr Val His Val Asn Gly Ala Lys Phe Ile Asp Thr Gln
295 300 305

AAT GGA TTA AGA ATC AAA ACA TGG CAG GGT GGT TCA GGC ATG GCA AGC
1013
Asn Gly Leu Arg Ile Lys Thr Trp Gln Gly Gly Ser Gly Met Ala Ser
310 315 320

CAT ATA ATT TAT GAG AAT GTT GAA ATG ATA AAT TCG GAG AAC CCC ATA
1061
His Ile Ile Tyr Glu Asn Val Glu Met Ile Asn Ser Glu Asn Pro Ile
325 330 335 340

TTA ATA AAT CAA TTC TAC TGC ACT TCA GCT TCT GCT TGC CAA AAC CAG
1109
Leu Ile Asn Gln Phe Tyr Cys Thr Ser Ala Ser Ala Cys Gln Asn Gln
345 350 355

AGG TCT GCG GTT CAA ATC CAA GAT GTG ACA TAC AAG AAC ATA CGT GGG
1157
Arg Ser Ala Val Gln Ile Gln Asp Val Thr Tyr Lys Asn Ile Arg Gly
360 365 370

ACA TCA GCA ACA GCA GCA GCA ATT CAA CTT AAG TGC AGT GAC AGT ATG
1205
Thr Ser Ala Thr Ala Ala Ala Ile Gln Leu Lys Cys Ser Asp Ser Met
375 380 385

CCC TGC AAA GAT ATA AAG CTA AGT GAT ATA TCT TTG AAG CTT ACC TCA
1253

```
Pro Cys Lys Asp Ile Lys Leu Ser Asp Ile Ser Leu Lys Leu Thr Ser
    390                 395                 400

GGG AAA ATT GCT TCC TGC CTT AAT GAT AAT GCA AAT GGA TAT TTC AGT
1301
Gly Lys Ile Ala Ser Cys Leu Asn Asp Asn Ala Asn Gly Tyr Phe Ser
405                 410                 415                 420

GGA CAC GTC ATC CCT GCA TGC AAG AAT TTA AGT CCA AGT GCT AAG CGA
1349
Gly His Val Ile Pro Ala Cys Lys Asn Leu Ser Pro Ser Ala Lys Arg
                425                 430                 435

AAA GAA TCT AAA TCC CAT AAA CAC CCA AAA ACT GTA ATG GTT GAA AAT
1397
Lys Glu Ser Lys Ser His Lys His Pro Lys Thr Val Met Val Glu Asn
            440                 445                 450

ATG CGA GCA TAT GAC AAG GGT AAC AGA ACA CGC ATA TTG TTG GGG TCG
1445
Met Arg Ala Tyr Asp Lys Gly Asn Arg Thr Arg Ile Leu Leu Gly Ser
        455                 460                 465

AGG CCT CCG AAT TGT ACA AAC AAA TGT CAT GGT TGC AGT CCA TGT AAG
1493
Arg Pro Pro Asn Cys Thr Asn Lys Cys His Gly Cys Ser Pro Cys Lys
    470                 475                 480

GCC AAG TTA GTT ATT GTT CAT CGT ATT ATG CCG CAG GAG TAT TAT CCT
1541
Ala Lys Leu Val Ile Val His Arg Ile Met Pro Gln Glu Tyr Tyr Pro
485                 490                 495                 500

CAG AGG TGG ATA TGC AGC TGT CAT GGC AAA ATC TAC CAT CCA TAATGAGATA
1593
Gln Arg Trp Ile Cys Ser Cys His Gly Lys Ile Tyr His Pro
                505                 510

CATTGAAACT GTATGTGCTA GTGAATATTC TTGTGGTACA ATATTAGAAC TGATATTGAA
1653

AATAAATCAT CAATGTTTCT AAGGCATTTA TAATAGATTA TATTAATGGT TCAGCCTGGT
1713

GCAAAAAAAA AAA
1726
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 514 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Met Lys Leu Ile Ala Pro Met Ala Phe Leu Ala Met Gln Leu
  1               5                  10                  15

Ile Ile Met Ala Ala Ala Glu Asp Gln Ser Ala Gln Ile Met Leu Asp
             20                  25                  30
```

Ser Val Val Glu Lys Tyr Leu Arg Ser Asn Arg Ser Leu Arg Lys Val
35 40 45

Glu His Ser Arg His Asp Ala Ile Asn Ile Phe Asn Val Glu Lys Tyr
50 55 60

Gly Ala Val Gly Asp Gly Lys His Asp Cys Thr Glu Ala Phe Ser Thr
65 70 75 80

Ala Trp Gln Ala Ala Cys Lys Asn Pro Ser Ala Met Leu Leu Val Pro
85 90 95

Gly Ser Lys Lys Phe Val Val Asn Asn Leu Phe Phe Asn Gly Pro Cys
100 105 110

Gln Pro His Phe Thr Phe Lys Val Asp Gly Ile Ile Ala Ala Tyr Gln
115 120 125

Asn Pro Ala Ser Trp Lys Asn Asn Arg Ile Trp Leu Gln Phe Ala Lys
130 135 140

Leu Thr Gly Phe Thr Leu Met Gly Lys Gly Val Ile Asp Gly Gln Gly
145 150 155 160

Lys Gln Trp Trp Ala Gly Gln Cys Lys Trp Val Asn Gly Arg Glu Ile
165 170 175

Cys Asn Asp Arg Asp Arg Pro Thr Ala Ile Lys Phe Asp Phe Ser Thr
180 185 190

Gly Leu Ile Ile Gln Gly Leu Lys Leu Met Asn Ser Pro Glu Phe His
195 200 205

Leu Val Phe Gly Asn Cys Glu Gly Val Lys Ile Ile Gly Ile Ser Ile
210 215 220

Thr Ala Pro Arg Asp Ser Pro Asn Thr Asp Gly Ile Asp Ile Phe Ala
225 230 235 240

Ser Lys Asn Phe His Leu Gln Lys Asn Thr Ile Gly Thr Gly Asp Asp
245 250 255

Cys Val Ala Ile Gly Thr Gly Ser Ser Asn Ile Val Ile Glu Asp Leu
260 265 270

Ile Cys Gly Pro Gly His Gly Ile Ser Ile Gly Ser Leu Gly Arg Glu
275 280 285

Asn Ser Arg Ala Glu Val Ser Tyr Val His Val Asn Gly Ala Lys Phe
290 295 300

Ile Asp Thr Gln Asn Gly Leu Arg Ile Lys Thr Trp Gln Gly Gly Ser
305 310 315 320

Gly Met Ala Ser His Ile Ile Tyr Glu Asn Val Glu Met Ile Asn Ser
325 330 335

Glu Asn Pro Ile Leu Ile Asn Gln Phe Tyr Cys Thr Ser Ala Ser Ala
340 345 350

Cys Gln Asn Gln Arg Ser Ala Val Gln Ile Gln Asp Val Thr Tyr Lys
355 360 365

Asn Ile Arg Gly Thr Ser Ala Thr Ala Ala Ala Ile Gln Leu Lys Cys
370 375 380

```
Ser Asp Ser Met Pro Cys Lys Asp Ile Lys Leu Ser Asp Ile Ser Leu
385                 390                 395                 400

Lys Leu Thr Ser Gly Lys Ile Ala Ser Cys Leu Asn Asp Asn Ala Asn
                405                 410                 415

Gly Tyr Phe Ser Gly His Val Ile Pro Ala Cys Lys Asn Leu Ser Pro
            420                 425                 430

Ser Ala Lys Arg Lys Glu Ser Lys Ser His Lys His Pro Lys Thr Val
        435                 440                 445

Met Val Glu Asn Met Arg Ala Tyr Asp Lys Gly Asn Arg Thr Arg Ile
    450                 455                 460

Leu Leu Gly Ser Arg Pro Pro Asn Cys Thr Asn Lys Cys His Gly Cys
465                 470                 475                 480

Ser Pro Cys Lys Ala Lys Leu Val Ile Val His Arg Ile Met Pro Gln
                485                 490                 495

Glu Tyr Tyr Pro Gln Arg Trp Ile Cys Ser Cys His Gly Lys Ile Tyr
            500                 505                 510

His Pro
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Arg Lys Val Glu His Ser Arg His Asp Ala Ile Asn Ile Phe Asn Val
1               5                   10                  15

Glu Lys Tyr Gly Ala Val Gly Asp Gly Lys His Asp Cys Thr Glu Ala
            20                  25                  30

Phe Ser Thr Ala Trp Gln Ala Ala Cys Lys Asn Pro Ser
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Arg Lys Val Glu His Ser Arg His Asp Ala Ile Asn Ile Phe Asn Val
1 5 10 15

Glu Lys Tyr Gly Ala Val Gly Asp Gly Lys His Asp Cys Thr Glu Ala
20 25 30

Phe Ser Thr Ala Trp Gln Lys Asn Pro
35 40

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Ser Arg His Asp Ala Ile Asn Ile Phe Asn Val Glu Lys Tyr Gly Ala
1 5 10 15

Val Gly Asp Gly Lys His Asp Cys Thr Glu Ala Phe Ser Thr Ala Trp
20 25 30

Gln Lys Asn Pro
35

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ala Ile Asn Ile Phe Asn Val Glu Lys Tyr
1 5 10

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1410 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGAAAAGTTG AGCATTCTCG TCATGATGCT ATCAACATCT TCAATGTGGA AAAGTATGGC
60

GCAGTAGGCG ATGGAAAGCA TGATTGCACT GAGGCATTTT CAACAGCATG GCAAGCTGCA
120

TGCAAAAACC CATCAGCAAT GTTGCTTGTG CCAGGCAGCA AGAAATTTGT TGTAAACAAT
180

CTGTTCTTCA ATGGGCCATG TCAACCTCAC TTTACTTTTA AGGTAGATGG GATAATAGCT
240

GCGTACCAAA ATCCAGCGAG CTGGAAGAAT AATAGAATAT GGTTGCAGTT TGCTAAACTT
300

ACAGGTTTTA CTCTAATGGG TAAAGGTGTA ATTGATGGGC AAGGAAAACA ATGGTGGGCT
360

GGCCAATGTA AATGGGTCAA TGGACGAGAA ATTTGCAACG ATCGTGATAG ACCAACAGCC
420

ATTAAATTCG ATTTTTCCAC GGGTCTGATA ATCCAAGGAC TGAAACTAAT GAACAGTCCC
480

GAATTTCATT TAGTTTTTGG GAATTGTGAG GGAGTAAAAA TCATCGGCAT TAGTATTACG
540

GCACCGAGAG ACAGTCCTAA CACTGATGGA ATTGATATCT TTGCATCTAA AAACTTTCAC
600

TTACAAAAGA ACACGATAGG AACAGGGGAT GACTGCGTCG CTATAGGCAC AGGGTCTTCT
660

AATATTGTGA TTGAGGATCT GATTTGCGGT CCAGGCCATG GAATAAGTAT AGGAAGTCTT
720

GGGAGGGAAA ACTCTAGAGC AGAGGTTTCA TACGTGCACG TAAATGGGGC TAAATTCATA
780

GACACACAAA ATGGATTAAG AATCAAAACA TGGCAGGGTG GTTCAGGCAT GGCAAGCCAT
840

ATAATTTATG AGAATGTTGA AATGATAAAT TCGGAGAACC CCATATTAAT AAATCAATTC
900

TACTGCACTT CAGCTTCTGC TTGCCAAAAC CAGAGGTCTG CGGTTCAAAT CCAAGATGTG
960

ACATACAAGA ACATACGTGG GACATCAGCA ACAGCAGCAG CAATTCAACT TAAGTGCAGT
1020

GACAGTATGC CCTGCAAAGA TATAAAGCTA AGTGATATAT CTTTGAAGCT TACCTCAGGG
1080

AAAATTGCTT CCTGCCTTAA TGATAATGCA AATGGATATT TCAGTGGACA CGTCATCCCT
1140

GCATGCAAGA ATTTAAGTCC AAGTGCTAAG CGAAAAGAAT CTAAATCCCA TAAACACCCA
1200

AAAACTGTAA TGGTTGAAAA TATGCGAGCA TATGACAAGG GTAACAGAAC ACGCATATTG
1260

TTGGGGTCGA GGCCTCCGAA TTGTACAAAC AAATGTCATG GTTGCAGTCC ATGTAAGGCC

1320

AAGTTAGTTA TTGTTCATCG TATTATGCCG CAGGAGTATT ATCCTCAGAG GTGGATATGC
1380

AGCTGTCATG GCAAAATCTA CCATCCATAA
1410

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1395 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCTCGTCATG ATGCTATCAA CATCTTCAAT GTGGAAAAGT ATGGCGCAGT AGGCGATGGA
60

AAGCATGATT GCACTGAGGC ATTTTCAACA GCATGGCAAG CTGCATGCAA AAACCCATCA
120

GCAATGTTGC TTGTGCCAGG CAGCAAGAAA TTTGTTGTAA ACAATCTGTT CTTCAATGGG
180

CCATGTCAAC CTCACTTTAC TTTTAAGGTA GATGGGATAA TAGCTGCGTA CCAAAATCCA
240

GCGAGCTGGA AGAATAATAG AATATGGTTG CAGTTTGCTA AACTTACAGG TTTTACTCTA
300

ATGGGTAAAG GTGTAATTGA TGGGCAAGGA AAACAATGGT GGGCTGGCCA ATGTAAATGG
360

GTCAATGGAC GAGAAATTTG CAACGATCGT GATAGACCAA CAGCCATTAA ATTCGATTTT
420

TCCACGGGTC TGATAATCCA AGGACTGAAA CTAATGAACA GTCCCGAATT TCATTTAGTT
480

TTTGGGAATT GTGAGGGAGT AAAAATCATC GGCATTAGTA TTACGGCACC GAGAGACAGT
540

CCTAACACTG ATGGAATTGA TATCTTTGCA TCTAAAAACT TTCACTTACA AAAGAACACG
600

ATAGGAACAG GGGATGACTG CGTCGCTATA GGCACAGGGT CTTCTAATAT TGTGATTGAG
660

GATCTGATTT GCGGTCCAGG CCATGGAATA AGTATAGGAA GTCTTGGGAG GGAAAACTCT
720

AGAGCAGAGG TTTCATACGT GCACGTAAAT GGGGCTAAAT TCATAGACAC ACAAAATGGA
780

TTAAGAATCA AAACATGGCA GGGTGGTTCA GGCATGGCAA GCCATATAAT TTATGAGAAT
840

GTTGAAATGA TAAATTCGGA GAACCCCATA TTAATAAATC AATTCTACTG CACTTCAGCT
900

TCTGCTTGCC AAAACCAGAG GTCTGCGGTT CAAATCCAAG ATGTGACATA CAAGAACATA
960

CGTGGGACAT CAGCAACAGC AGCAGCAATT CAACTTAAGT GCAGTGACAG TATGCCCTGC
1020

AAAGATATAA AGCTAAGTGA TATATCTTTG AAGCTTACCT CAGGGAAAAT TGCTTCCTGC
1080

CTTAATGATA ATGCAAATGG ATATTTCAGT GGACACGTCA TCCCTGCATG CAAGAATTTA
1140

AGTCCAAGTG CTAAGCGAAA AGAATCTAAA TCCCATAAAC ACCCAAAAAC TGTAATGGTT
1200

GAAAATATGC GAGCATATGA CAAGGGTAAC AGAACACGCA TATTGTTGGG GTCGAGGCCT
1260

CCGAATTGTA CAAACAAATG TCATGGTTGC AGTCCATGTA AGGCCAAGTT AGTTATTGTT
1320

CATCGTATTA TGCCGCAGGA GTATTATCCT CAGAGGTGGA TATGCAGCTG TCATGGCAAA
1380

ATCTACCATC CATAA
1395

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1479 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAAGATCAAT CTGCCCAAAT TATGTTGGAC AGTGTTGTCG AAAAATATCT TAGATCGAAT
60

CGGAGTTTAA GAAAAGTTGA GCATTCTCGT CATGATGCTA TCAACATCTT CAATGTGGAA
120

AAGTATGGCG CAGTAGGCGA TGGAAAGCAT GATTGCACTG AGGCATTTTC AACAGCATGG
180

CAAGCTGCAT GCAAAAACCC ATCAGCAATG TTGCTTGTGC CAGGCAGCAA GAAATTTGTT
240

GTAAACAATC TGTTCTTCAA TGGGCCATGT CAACCTCACT TTACTTTTAA GGTAGATGGG
300

ATAATAGCTG CGTACCAAAA TCCAGCGAGC TGGAAGAATA ATAGAATATG GTTGCAGTTT
360

GCTAAACTTA CAGGTTTTAC TCTAATGGGT AAAGGTGTAA TTGATGGGCA AGGAAAACAA
420

TGGTGGGCTG GCCAATGTAA ATGGGTCAAT GGACGAGAAA TTTGCAACGA TCGTGATAGA
480

CCAACAGCCA TTAAATTCGA TTTTTCCACG GGTCTGATAA TCCAAGGACT GAAACTAATG
540

AACAGTCCCG AATTTCATTT AGTTTTTGGG AATTGTGAGG GAGTAAAAAT CATCGGCATT
600

AGTATTACGG CACCGAGAGA CAGTCCTAAC ACTGATGGAA TTGATATCTT TGCATCTAAA
660

AACTTTCACT TACAAAAGAA CACGATAGGA ACAGGGGATG ACTGCGTCGC TATAGGCACA
720

GGGTCTTCTA ATATTGTGAT TGAGGATCTG ATTTGCGGTC CAGGCCATGG AATAAGTATA
780

GGAAGTCTTG GGAGGGAAAA CTCTAGAGCA GAGGTTTCAT ACGTGCACGT AAATGGGGCT
840

AAATTCATAG ACACACAAAA TGGATTAAGA ATCAAAACAT GGCAGGGTGG TTCAGGCATG
900

GCAAGCCATA TAATTTATGA GAATGTTGAA ATGATAAATT CGGAGAACCC CATATTAATA
960

AATCAATTCT ACTGCACTTC AGCTTCTGCT TGCCAAAACC AGAGGTCTGC GGTTCAAATC
1020

CAAGATGTGA CATACAAGAA CATACGTGGG ACATCAGCAA CAGCAGCAGC AATTCAACTT
1080

AAGTGCAGTG ACAGTATGCC CTGCAAAGAT ATAAAGCTAA GTGATATATC TTTGAAGCTT
1140

ACCTCAGGGA AAATTGCTTC CTGCCTTAAT GATAATGCAA ATGGATATTT CAGTGGACAC
1200

GTCATCCCTG CATGCAAGAA TTTAAGTCCA AGTGCTAAGC GAAAAGAATC TAAATCCCAT
1260

AAACACCCAA AAACTGTAAT GGTTGAAAAT ATGCGAGCAT ATGACAAGGG TAACAGAACA
1320

CGCATATTGT TGGGGTCGAG GCCTCCGAAT TGTACAAACA AATGTCATGG TTGCAGTCCA
1380

TGTAAGGCCA AGTTAGTTAT TGTTCATCGT ATTATGCCGC AGGAGTATTA TCCTCAGAGG
1440

TGGATATGCA GCTGTCATGG CAAAATCTAC CATCCATAA
1479

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGGTCTAGAG GTACCGTCCG TCCGATCGAT CCATT
35

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AATGATCGAT GCT
13

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

RTAYTTYTCN ACRTTRAA
18

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGGTCTAGAG GTA
13

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Phe Asn Val Glu Lys Tyr
1 5

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCTGCAGTAY TTYTCNACRT TRAANAT
27

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCTGCAG
7

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ile Phe Asn Val Glu Lys Tyr
1 5

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCTGCAGTAY TTYTCNACRT TRAADAT
27

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCNATHAAYA THTTYAA
17

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Ala Ile Asn Ile Phe Asn
1 5

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGAATTCCGC NATHAAYATH TTYAAYGT
28

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGAATTCC
8

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Ala Ile Asn Ile Phe Asn Val
1 5

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCYTCNGTRC ARTCRTGYTT
20

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Lys His Asp Cys Thr Glu Ala
1 5

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGCTGCAGGT RCARTCRTGY TTNCCRTC
28

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGCTGCAG
8

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Asp Gly Lys His Asp Cys Thr
1 5

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

ATGTTGGACA GTGTTGTCGA A
21

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GGGAATTCAG AAAAGTTGAG CATTCTCGT
29

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GGGAATTC
8

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GTTCTTCAAT GGGCCATGT
19

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GTGTTAGGAC TGTCTCTCGG
20

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TGTCCAGGCC ATGGAATAAG
20

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GCCTTACATG GACTGCAACC
20

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TCCACGGGTC TGATAATCCA
20

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AGGCAGGAAG CAATTTTCCC
20

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TACTGCACTT CAGCTTCTGC
20

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GGGGGTCTCC GAATTTATCA
20

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGATATTTCA GTGGACACGT
20

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

TATTAGAAGA CCCTGCGCCT
20

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CCATGTAAGG CCAAGTTAGT
20

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

ACACCTTTAC CCATTAGAGT
20

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CTGTCCAACA TAATTTGGGC
20

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CATGGCAGGG TGGTTCAGGC
20

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TAGCCCCATT TACGTGCACG
20

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TTGGGGTCGA GGCCTCCGAA
20

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

TAAAAUGGC
9

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AACAAUGGC
9

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GCCGAATTCA TGGCCATGAA ATTAATT
27

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GCCGAATTC
9

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CGGGGATCCT CATTATGGAT GGTAGAT
27

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CGGGGATCC
9

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Phe Thr Phe Lys Val Asp Gly Ile Ile Ala Ala Tyr Gln

1 5 10

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Asn Gly Tyr Phe Ser Gly His Val Ile Pro Ala Cys Lys Asn
1 5 10

Claims:

1. A nucleic acid having a nucleotide sequence coding for a Japanese Cedar pollen allergen *Cry j* II, or at least one antigenic fragment thereof, or the functional equivalent of said nucleotide sequence.

2. A nucleic acid of claim 1 wherein said nucleotide sequence consists essentially of at least one fragment of the coding portion of the nucleotide sequence of Fig. 4 (SEQ ID NO: 1).

3. A nucleic acid of claim 2 wherein said fragment comprises bases 108 through 1586 (SEQ ID NO: 9) of the nucleotide sequence of Fig. 4 (SEQ ID NO: 1).

4. A nucleic acid of claim 1 wherein said nucleotide sequence consists essentially of the nucleotide sequence of Fig. 4 (SEQ ID NO: 1).

5. A nucleic acid of claim 1 wherein said fragment comprises bases selected from the group consisting of bases 177 through 1586 (SEQ ID NO: 7) of the nucleotide sequence of Fig. 4, and bases 192 through 1586 (SEQ ID NO: 8) of the nucleotide sequence of Fig. 4 (SEQ ID NO: 1).

6. An expression vector comprising a nucleotide sequence coding for a Japanese cedar pollen allergen *Cry j* II, or at least one antigenic fragment thereof, or the functional equivalent of said nucleotide sequence.

7. An expression vector of claim 6 wherein said nucleotide sequence consists essentially of at least one fragment of the coding portion of the nucleotide sequence of Fig. 4 (SEQ ID NO: 1).

8. An expression vector of claim 6 wherein said nucleotide sequence comprises bases 108 through 1586 (SEQ ID NO: 9) of the nucleotide sequence of Fig. 4.

9. A host cell transformed to express a protein or peptide encoded by the nucleic acid of claim 1.

10. Isolated *Cry j* II protein, or at least one antigenic fragment thereof, produced in a host cell transformed with the nucleic acid of claim 1.

11. An antigenic fragment of claim 10 which does not bind immunoglobulin E specific for a Japanese cedar pollen allergen, or if binding of said antigenic fragment to said immunoglobulin E occurs, such binding does not result in histamine release from mast cells or basophils.

12. An antigenic fragment of claim 10 which binds immunoglobulin E to a substantially lesser extent than purified, native *Cry j* II protein binds said immunoglobulin E.

13. Isolated *Cry j* II protein of claim 10 wherein the host cell is E.coli.

14. A method of producing *Cry j* II protein, or at least one fragment thereof, comprising the steps of:
    a. culturing a host cell transformed with a DNA sequence encoding *Cry j* II protein or fragment thereof, in an appropriate medium to produce a mixture of cells and medium containing *Cry j* II protein or at least one fragment thereof; and
    b. purifying said mixture to produce substantially pure *Cry j* II protein, or at least one fragment thereof.

15. A protein preparation comprising *Cry j* II protein, or at least one fragment thereof, synthesized in a host cell transformed with a nucleic acid comprising a nucleotide sequence encoding all or a portion of *Cry j* II.

16. A protein preparation of claim 15 wherein said at least one fragment of *Cry j* II is an antigenic fragment.

17. A protein preparation comprising chemically synthesized *Cry j* II protein, or at least one fragment thereof.

18. A protein preparation of claim 15 wherein said *Cry j* II protein comprises an amino acid sequence shown in Fig. 4 (SEQ ID NO: 2).

19. A protein preparation of claim 17 wherein said *Cry j* II protein comprises an amino acid sequence shown in Fig. 4 (SEQ ID NO: 2).

20. An isolated peptide comprising at least one T cell epitope of *Cry j* II.

21. An isolated peptide of claim 20 which as minimal immunoglobulin E stimulating activity.

22. An isolated peptide of claim 20 which does not bind immunoglobulin E specific for a Japanese cedar pollen allergen, or if binding of the peptide to said immunoglobulin E occurs, such binding does not result in histamine release from mast cells or basophils.

23. An isolated peptide of claim 20 which binds immunoglobulin E to a substantially lesser extent than purified native *Cry j* II protein binds said immunoglobulin E.

24. Isolated *Cry j* II protein, or an antigenic fragment thereof, which modifies, in an individual sensitive to Japanese cedar pollen to whom it is administered, the allergic response of the individual to a Japanese cedar pollen allergen.

25. Isolated *Cry j* II protein or antigenic fragment of claim 24 which modifies B-cell response of the individual to a Japanese cedar pollen allergen, T-cell response of the individual to a Japanese cedar pollen allergen, or both the B-cell response and the T-cell response of the individual to a Japanese cedar pollen allergen.

26. Modified *Cry j* II protein or at least one modified fragment thereof, which when administered to an individual sensitive to Japanese cedar pollen, reduces the allergic response of the individual to *Cry j* II.

27. A therapeutic composition comprising isolated *Cry j* II protein, or at least one fragment thereof, and a pharmaceutically acceptable carrier or diluent.

28. A therapeutic composition of claim 27 wherein said *Cry j* II protein comprises an amino acid sequence shown in Fig. 4 (SEQ ID NO: 1).

29. A method of treating sensitivity to a Japanese cedar pollen allergen, or an allergen immunologically cross-reactive with a Japanese cedar pollen allergen, in an individual sensitive to said allergen, comprising administering to the individual a therapeutically effective amount of the composition of claim 27.

30. A method of detecting sensitivity in an individual to a Japanese cedar pollen allergen, comprising combining a blood sample obtained from the individual with isolated *Cry j* II protein, or antigenic fragment thereof, produced in a host cell transformed with the nucleic acid of claim 1 or chemically synthesized, under conditions appropriate for binding of blood components with the protein or fragment thereof, and determining the extent to which such binding occurs.

31. A method of claim 30 wherein the extent to which binding occurs is determined by assessing T cell function, T cell proliferation, B cell function, binding of the protein or fragment thereof to antibodies present in the blood or a combination thereof.

32. A monoclonal antibody, polyclonal antibody or immunoreactive fragment thereof, specifically reactive with *Cry j* II protein, or at least one antigenic fragment thereof.

33. *Cry j* II protein isolated from Japanese cedar pollen, said protein having a molecular weight of about 40 kD as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

34. A host cell transformed with a vector containing the cDNA insert of *Cry j* II, said host cell having ATCC deposit number 69105.

35. A recombinant DNA molecule comprising a DNA coding for a polypeptide having at least one epitope of the protein allergen, *Cry j* II.

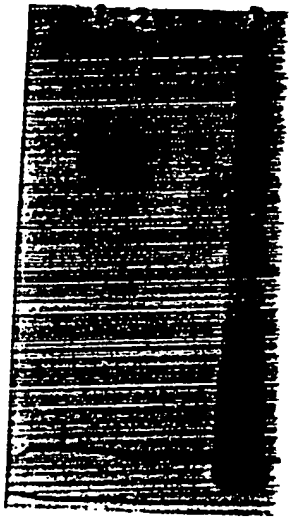

Fig. 1a

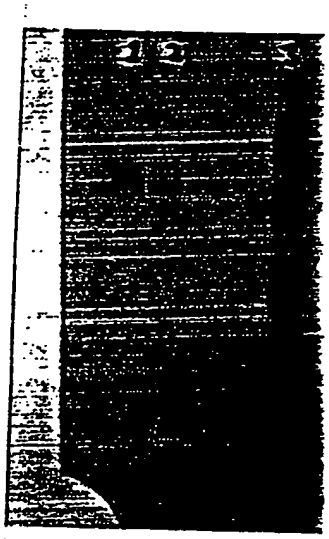

Fig. 1b

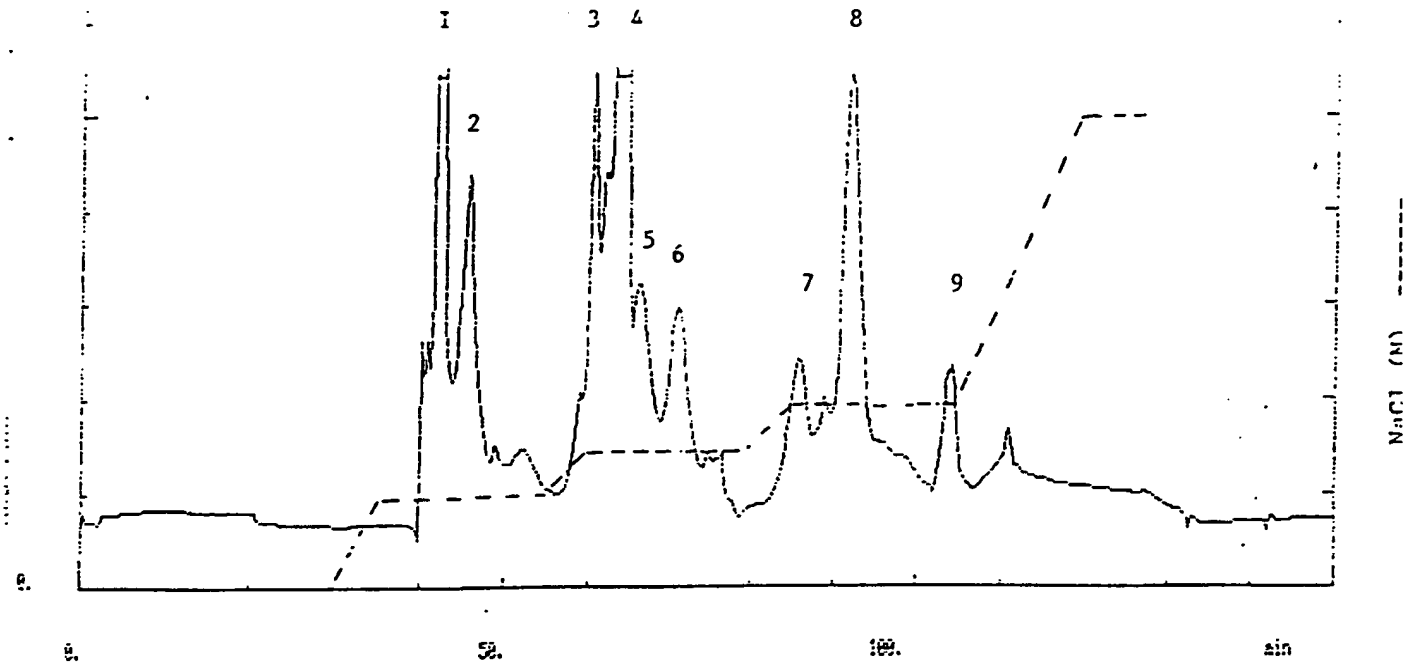

Fig. 2

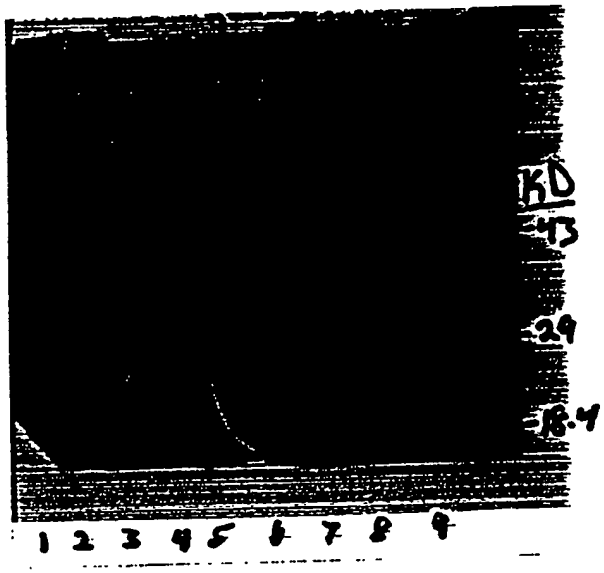

Fig 3

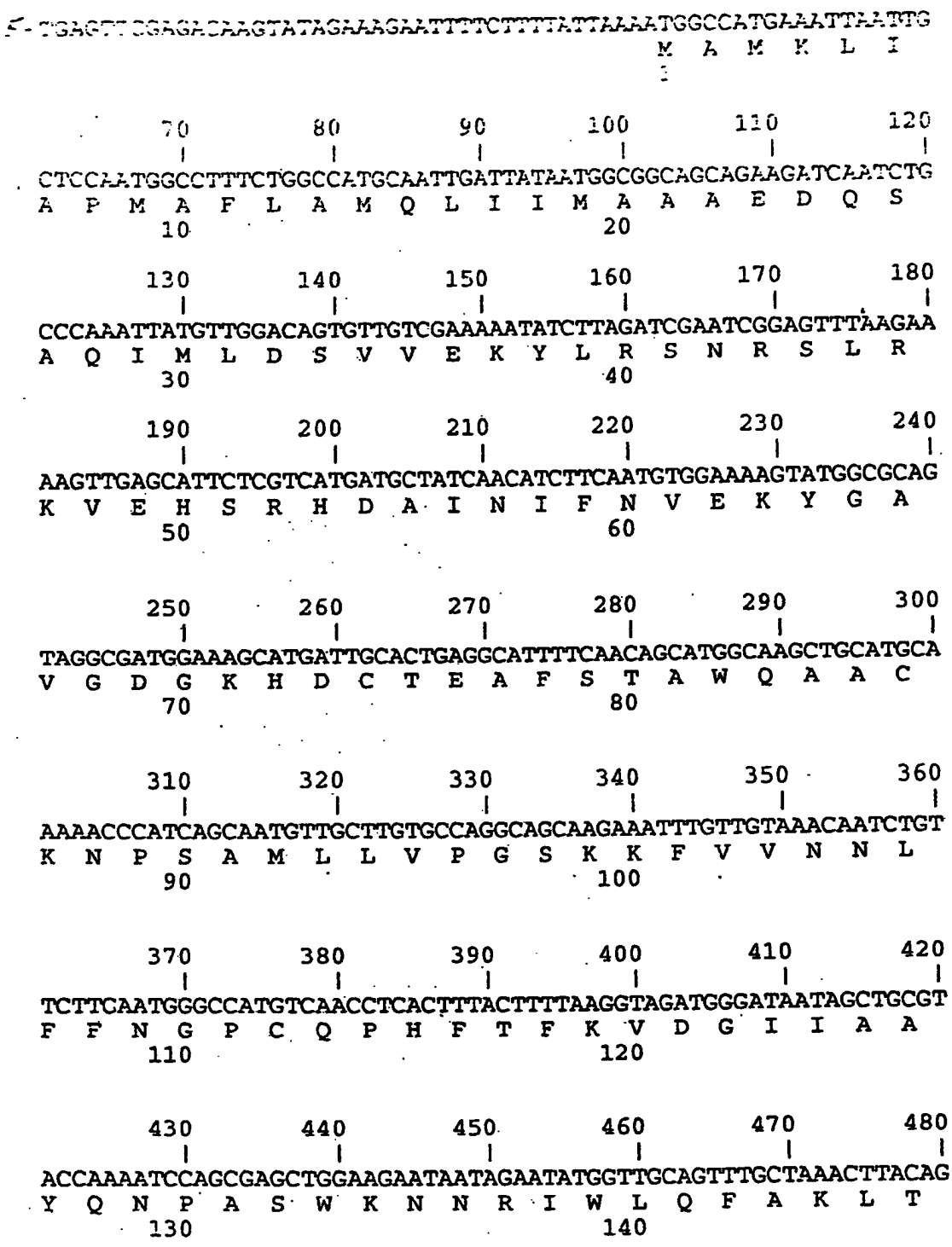
5'-TGAGTTCGAGACAAGTATAGAAAGAATTTTCTTTTATTAAAATGGCCATGAAATTAATTG
M A M K L I
70 80 90 100 110 120
CTCCAATGGCCTTTCTGGCCATGCAATTGATTATAATGGCGGCAGCAGAAGATCAATCTG
A P M A F L A M Q L I I M A A A E D Q S
10 20
130 140 150 160 170 180
CCCAAATTATGTTGGACAGTGTTGTCGAAAAATATCTTAGATCGAATCGGAGTTTAAGAA
A Q I M L D S V V E K Y L R S N R S L R
30 40
190 200 210 220 230 240
AAGTTGAGCATTCTCGTCATGATGCTATCAACATCTTCAATGTGGAAAAGTATGGCGCAG
K V E H S R H D A I N I F N V E K Y G A
50 60
250 260 270 280 290 300
TAGGCGATGGAAAGCATGATTGCACTGAGGCATTTTCAACAGCATGGCAAGCTGCATGCA
V G D G K H D C T E A F S T A W Q A A C
70 80
310 320 330 340 350 360
AAAACCCATCAGCAATGTTGCTTGTGCCAGGCAGCAAGAAATTTGTTGTAAACAATCTGT
K N P S A M L L V P G S K K F V V N N L
90 100
370 380 390 400 410 420
TCTTCAATGGGCCATGTCAACCTCACTTTACTTTTAAGGTAGATGGGATAATAGCTGCGT
F F N G P C Q P H F T F K V D G I I A A
110 120
430 440 450 460 470 480
ACCAAAATCCAGCGAGCTGGAAGAATAATAGAATATGGTTGCAGTTTGCTAAACTTACAG
Y Q N P A S W K N N R I W L Q F A K L T
130 140

Fig 7 cont.

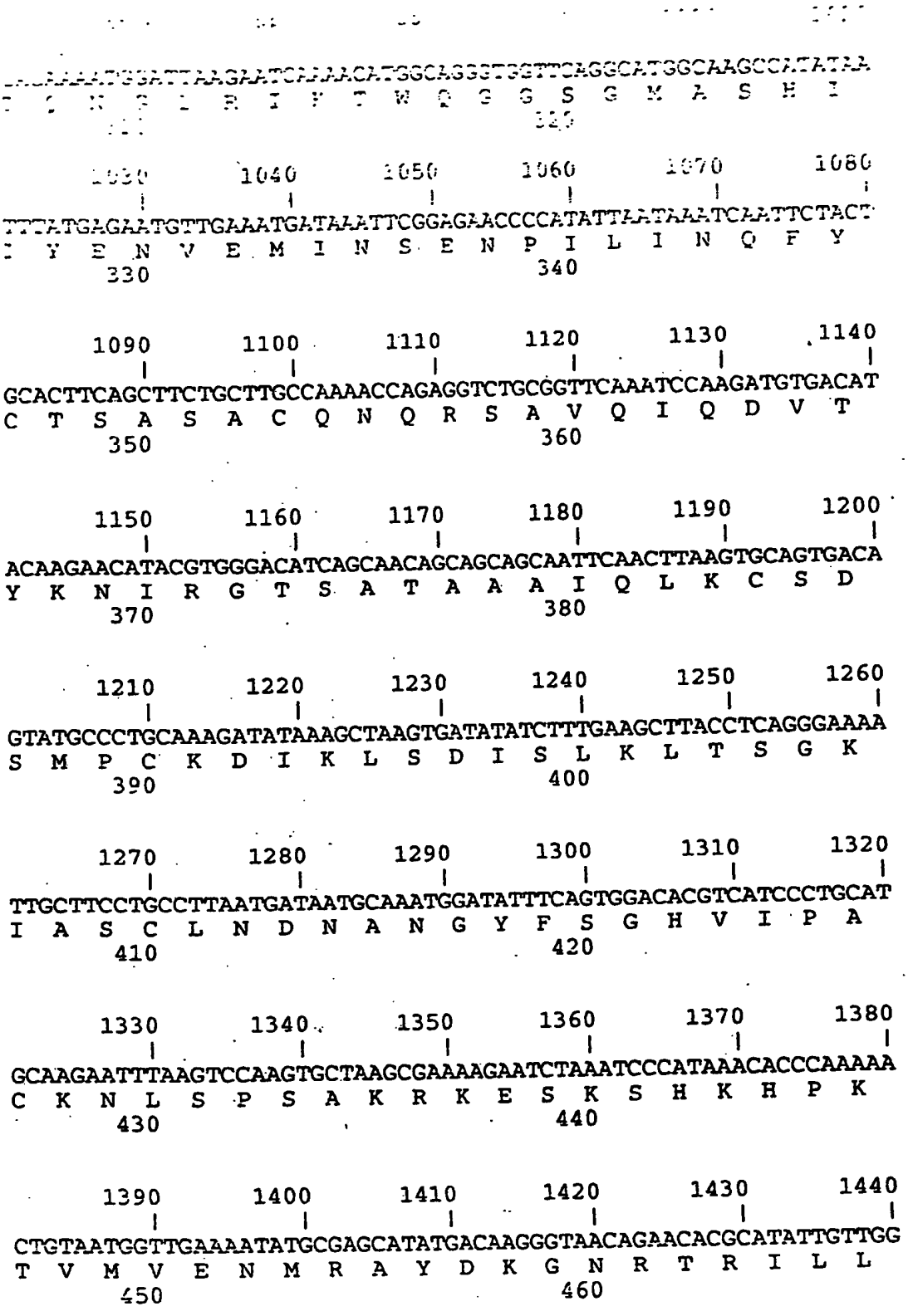
1030 1040 1050 1060 1070 1080
TTTATGAGAATGTTGAAATGATAAATTCGGAGAACCCCATATTAATAAATCAATTCTACT
I Y E N V E M I N S E N P I L I N Q F Y
330 340
1090 1100 1110 1120 1130 1140
GCACTTCAGCTTCTGCTTGCCAAAACCAGAGGTCTGCGGTTCAAATCCAAGATGTGACAT
C T S A S A C Q N Q R S A V Q I Q D V T
350 360
1150 1160 1170 1180 1190 1200
ACAAGAACATACGTGGGACATCAGCAACAGCAGCAGCAATTCAACTTAAGTGCAGTGACA
Y K N I R G T S A T A A A I Q L K C S D
370 380
1210 1220 1230 1240 1250 1260
GTATGCCCTGCAAAGATATAAAGCTAAGTGATATATCTTTGAAGCTTACCTCAGGGAAAA
S M P C K D I K L S D I S L K L T S G K
390 400
1270 1280 1290 1300 1310 1320
TTGCTTCCTGCCTTAATGATAATGCAAATGGATATTTCAGTGGACACGTCATCCCTGCAT
I A S C L N D N A N G Y F S G H V I P A
410 420
1330 1340 1350 1360 1370 1380
GCAAGAATTTAAGTCCAAGTGCTAAGCGAAAAGAATCTAAATCCCATAAACACCCAAAAA
C K N L S P S A K R K E S K S H K H P K
430 440
1390 1400 1410 1420 1430 1440
CTGTAATGGTTGAAAATATGCGAGCATATGACAAGGGTAACAGAACACGCATATTGTTGG
T V M V E N M R A Y D K G N R T R I L L
450 460

Fig 7 cont

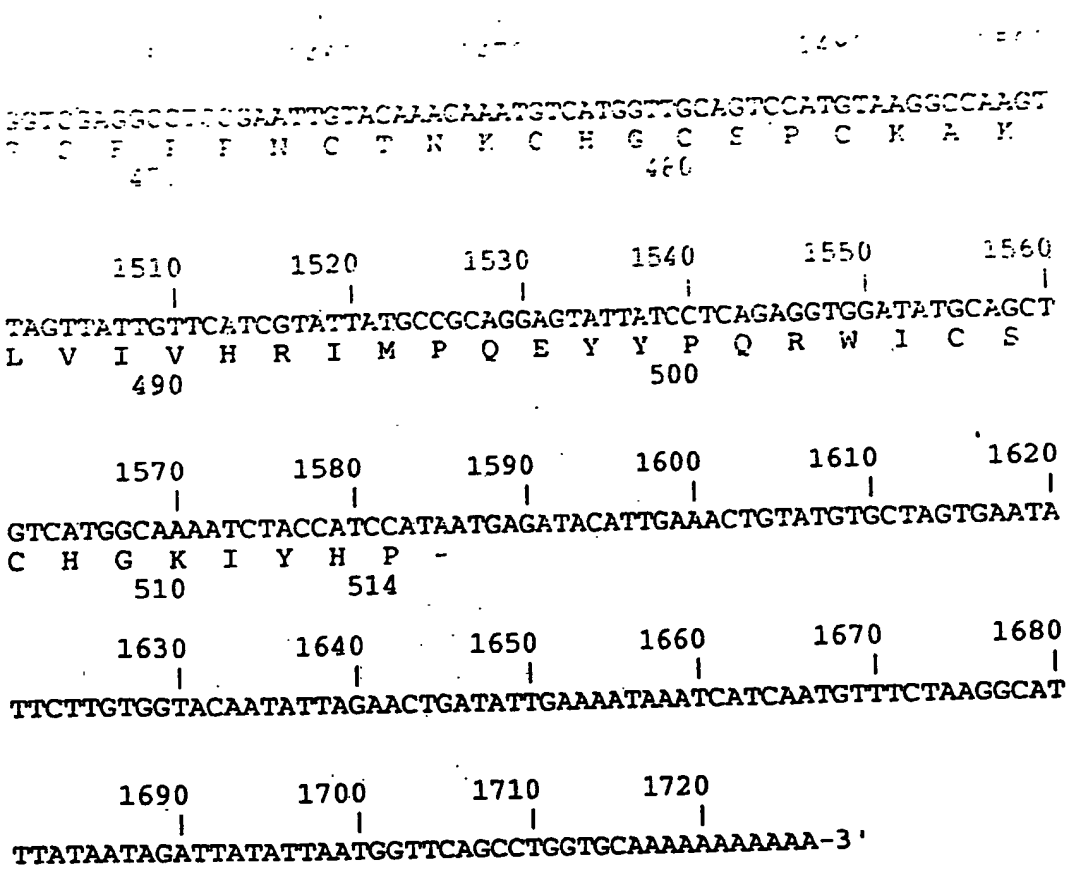
1510 1520 1530 1540 1550 1560
TAGTTATTGTTCATCGTATTATGCCGCAGGAGTATTATCCTCAGAGGTGGATATGCAGCT
L V I V H R I M P Q E Y Y P Q R W I C S
490 500
1570 1580 1590 1600 1610 1620
GTCATGGCAAAATCTACCATCCATAATGAGATACATTGAAACTGTATGTGCTAGTGAATA
C H G K I Y H P -
510 514
1630 1640 1650 1660 1670 1680
TTCTTGTGGTACAATATTAGAACTGATATTGAAAATAAATCATCAATGTTTCTAAGGCAT
1690 1700 1710 1720
TTATAATAGATTATATTAATGGTTCAGCCTGGTGCAAAAAAAAAAA-3'

Figure 5

```
          10        20        30        40
          |         |         .         |
MAMKLIAPMAFLAMQLIIMAAAEDQSAQIMLDSVVEKYLR

          50        60        70        80
          |         |         .         |
SNRSLRKVEHSRHDAINIFNVEKYGAVGDGKHDCTEAFST

          90        100       110       120
          |         |         |         |
AWQAACKNPSAMLLVPGSKKFVVNNLFFNGPCQPHFTFKV

          130       140       150       160
          |         ||        |         |
DGIIAAYQNPASWKNNRIWLQFAKLTGFTLMGKGVIDGQG

          170       180       190       200
          |         |         |         |
KQWWAGQCKWVNGREICNDRDRPTAIKFDFSTGLIIQGLK

          210       220       230       240
          |         |         |         |
LMNSPEFHLVFGNCEGVKIIGISITAPRDSPNTDGIDIFA

          250       260       270       280
          |         |         |         |
SKNFHLQKNTIGTGDDCVAIGTGSSNIVIEDLICGPGHGI

          290       300       310       320
          |         |         |         |
SIGSLGRENSRAEVSYVHVNGAKFIDTQNGLRIKTWQGGS

          330       340       350       360
          |         |         |         |
GMASHIIYENVEMINSENPILINQFYCTSASACQNQRSAV

          370       380       390       400
          |         |         |         |
QIQDVTYKNIRGTSATAAAIQLKCSDSMPCKDIKLSDISL

          410       420       430       440
          |         |         |         |
KLTSGKIASCLNDNANGYFSGHVIPACKNLSPSAKRKESK

          450       460       470       480
          |         |         |         |
SHKHPKTVMVENMRAYDKGNRTRILLGSRPPNCTNKCHGC

          490       500       510 514
          |         |         |   |
SPCKAKLVIVHRIMPQEYYPQRWICSCHGKIYHP
```

	50												60						
Cry j II	R	K	V	E	H	S	R	H	D	A	I	N	I	F	N	V	E	K	Y	G	A
Long	R	K	V	E	H	S	R	H	D	A	I	N	I	F	N	V	E	K	Y	G	A
Short						S	R	H	D	A	I	N	I	F	N	V	E	K	Y	G	A
Sakaguchi										A	I	N	I	F	N	V	E	K	Y		

	70											80									90		
Cry j II	V	G	D	G	K	H	D	C	T	E	A	F	S	T	A	W	Q	A	A	C	K	N	P	S
Long	V	G	D	G	K	H	D	C	T	E	A	F	S	T	A	W	(	Q	)	K	N	P	(	)
Short	V	G	D	G	K	H	D	C	T	E	A	F	S	T	A	W	(	Q	)	K	N	P	(	)

Fig. 6

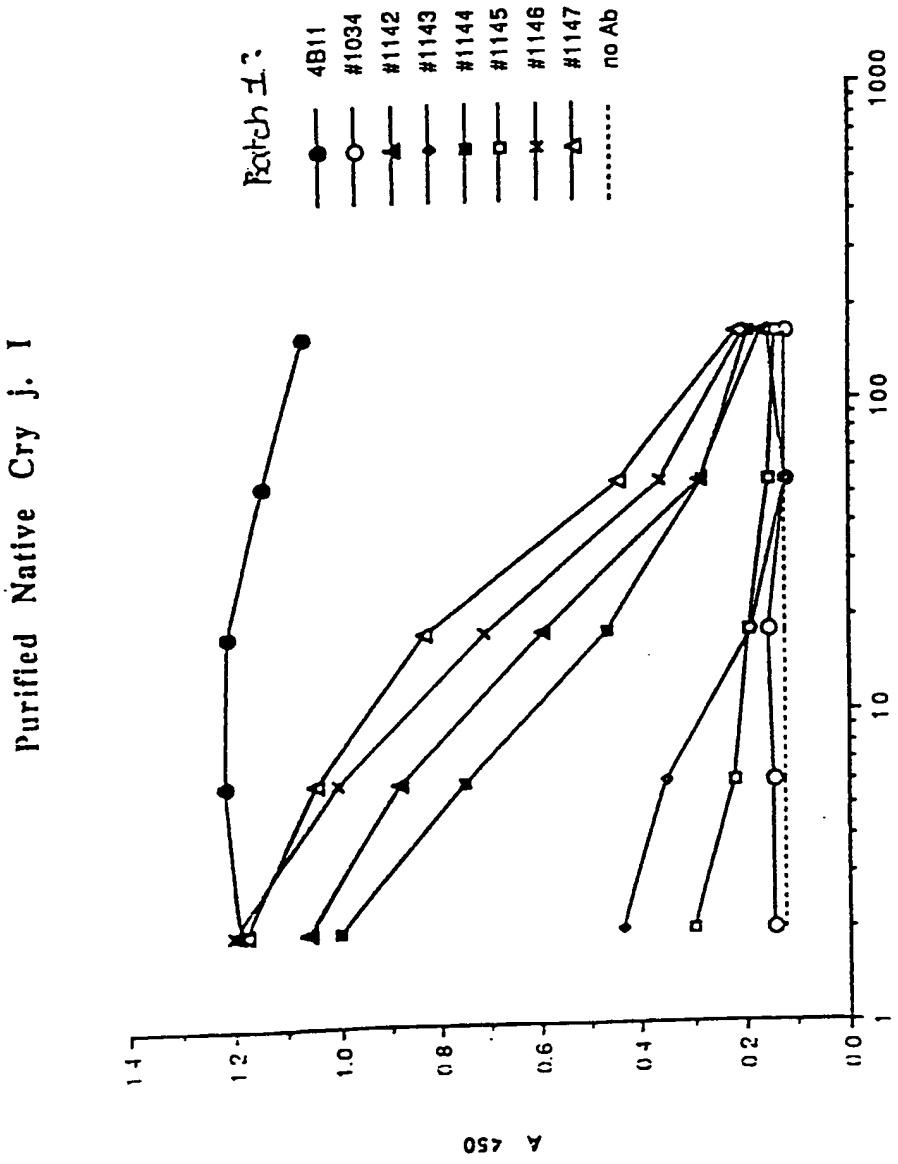
Purified Native Cry j. I
Batch 1?
4B11
#1034
#1142
#1143
#1144
#1145
#1146
#1147
no Ab
A 450
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1
10
100
1000
1/ dilution
(x1000 for 4B11)

Fig. 7

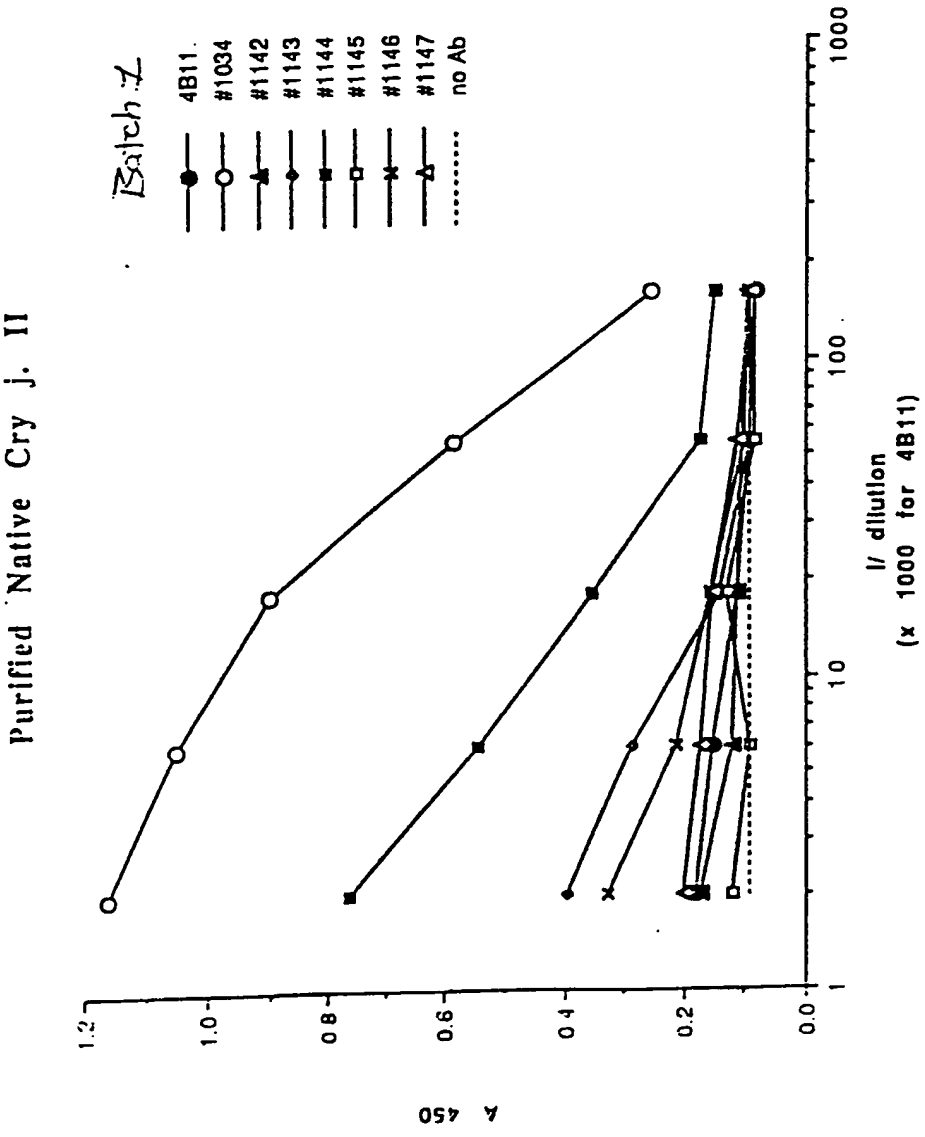
Purified Native Cry j. II
Batch 1
4B11
#1034
#1142
#1143
#1144
#1145
#1146
#1147
no Ab
A 450
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
10
100
1000
1/ dilution
(x 1000 for 4B11)

Fig.8

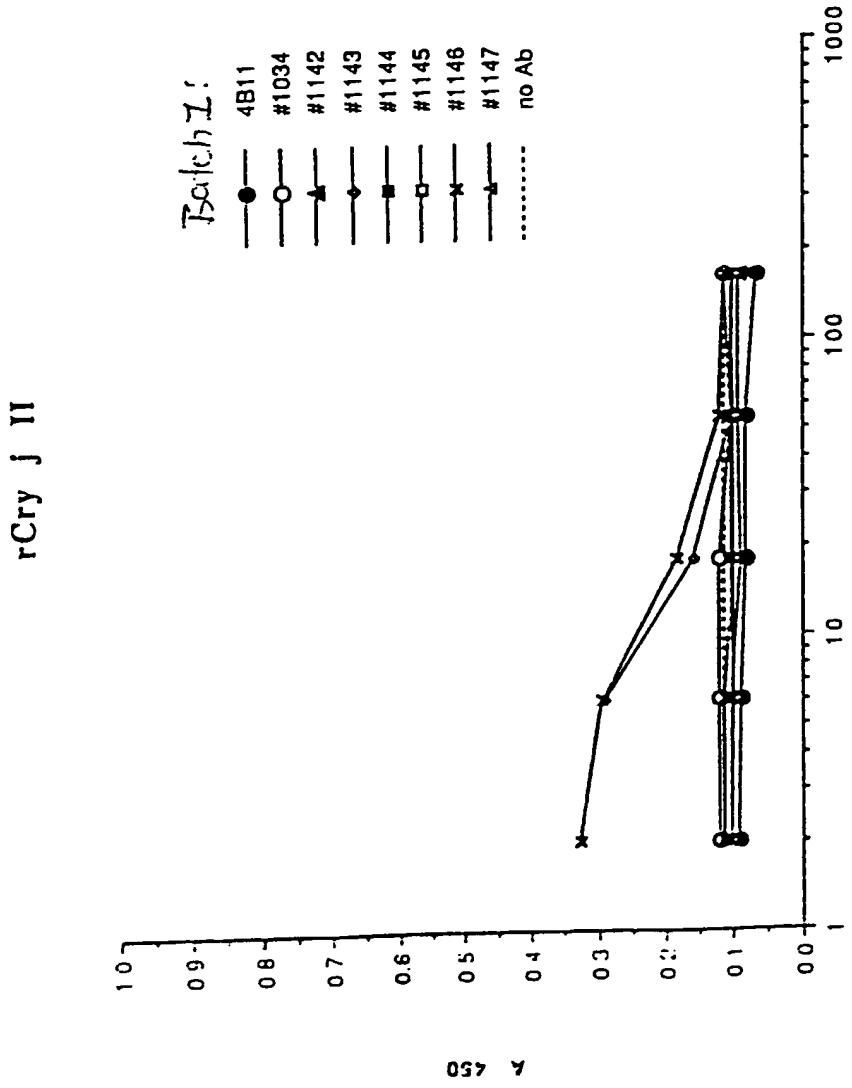
rCry j II
Batch 1:
4B11
#1034
#1142
#1143
#1144
#1145
#1146
#1147
no Ab
A 450
1/ dilution
(x 1000 for 4B11)

Fig. 9

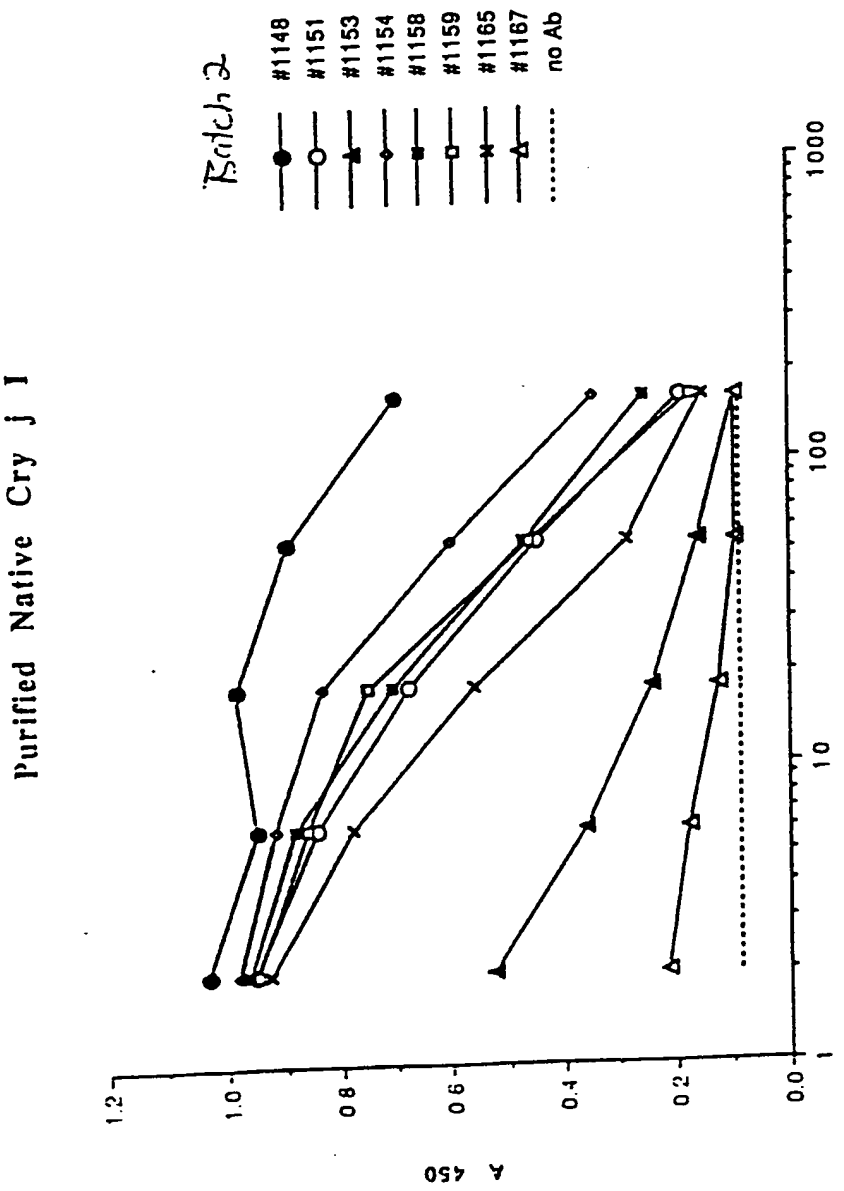
Purified Native Cry j I
Batch 2
#1148
#1151
#1153
#1154
#1158
#1159
#1165
#1167
no Ab
A 450
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
10
100
1000
1/ dilution

Fig. 10

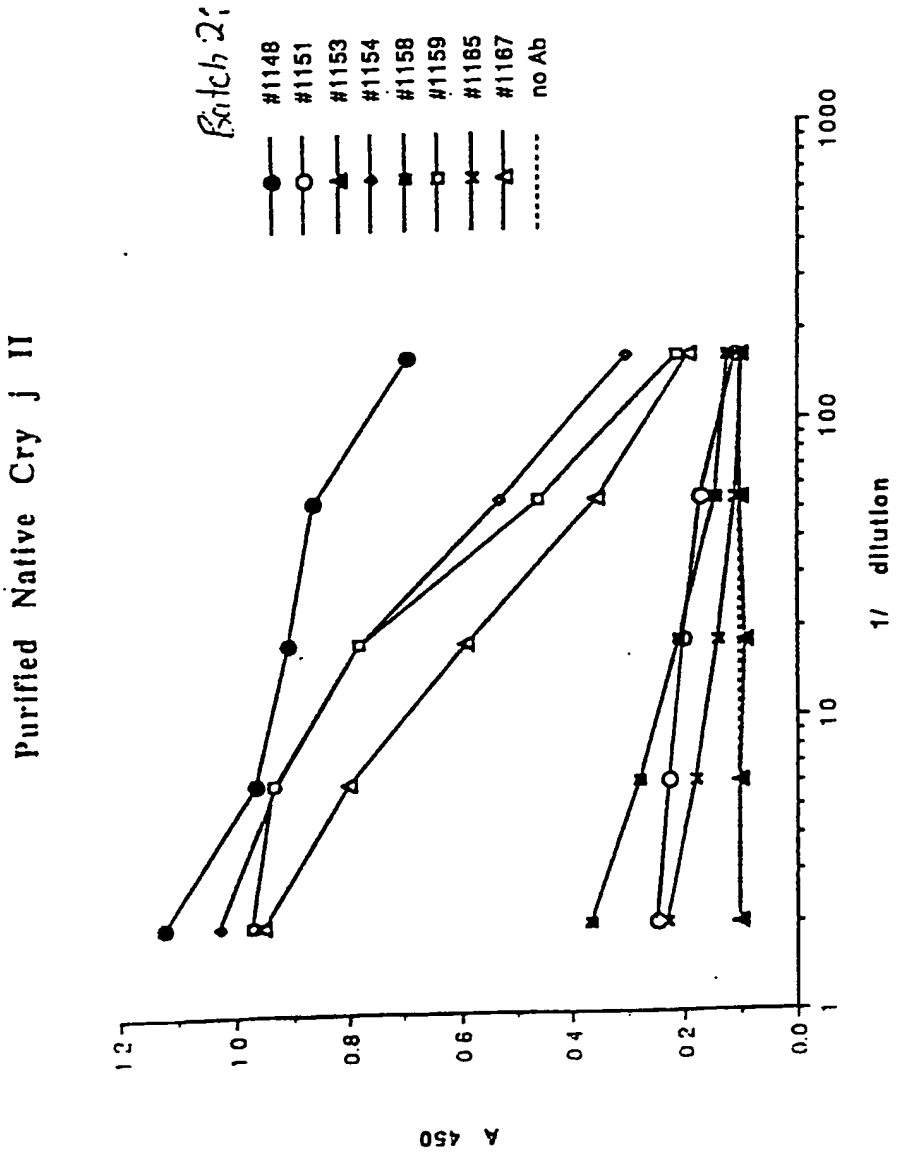
Purified Native Cry j II
Batch 2
#1148
#1151
#1153
#1154
#1158
#1159
#1165
#1167
no Ab
A 450
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
10
100
1000
1/ dilution

Fig. 11

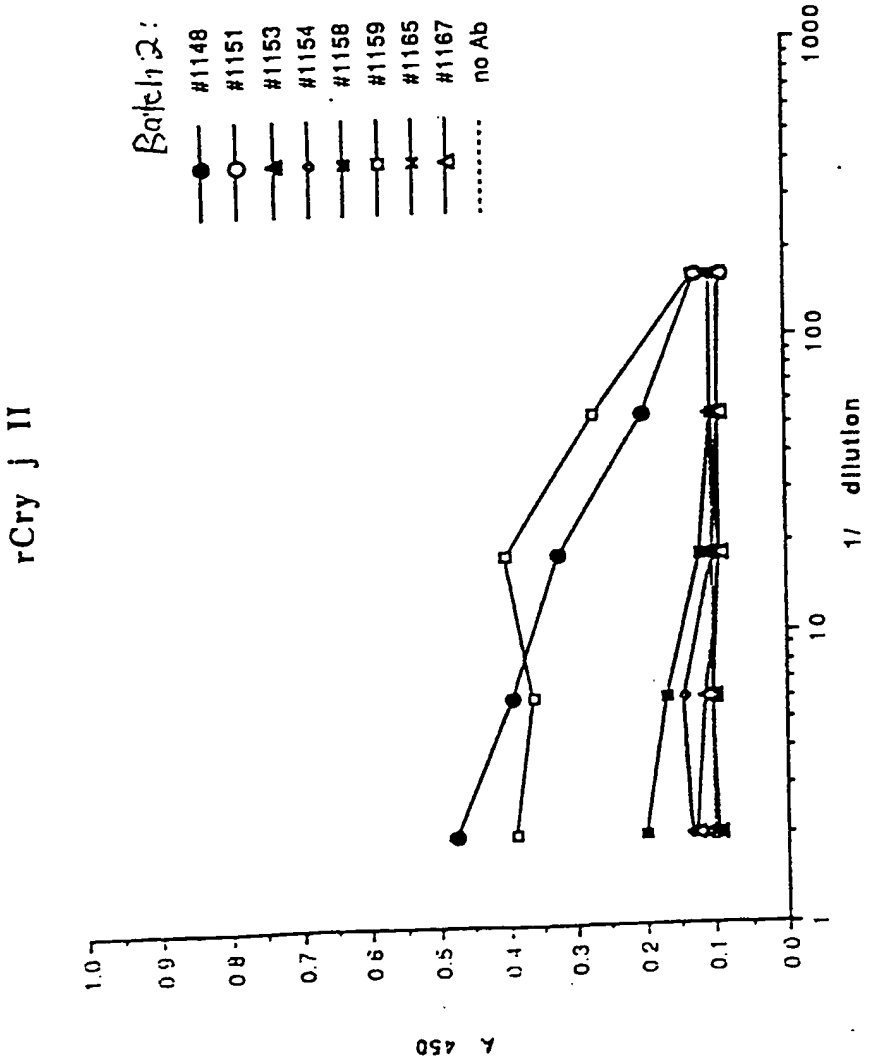
rCry j II
Batch 2:
#1148
#1151
#1153
#1154
#1158
#1159
#1165
#1167
no Ab
A 450
1/ dilution
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1
10
100
1000

Fig. 12

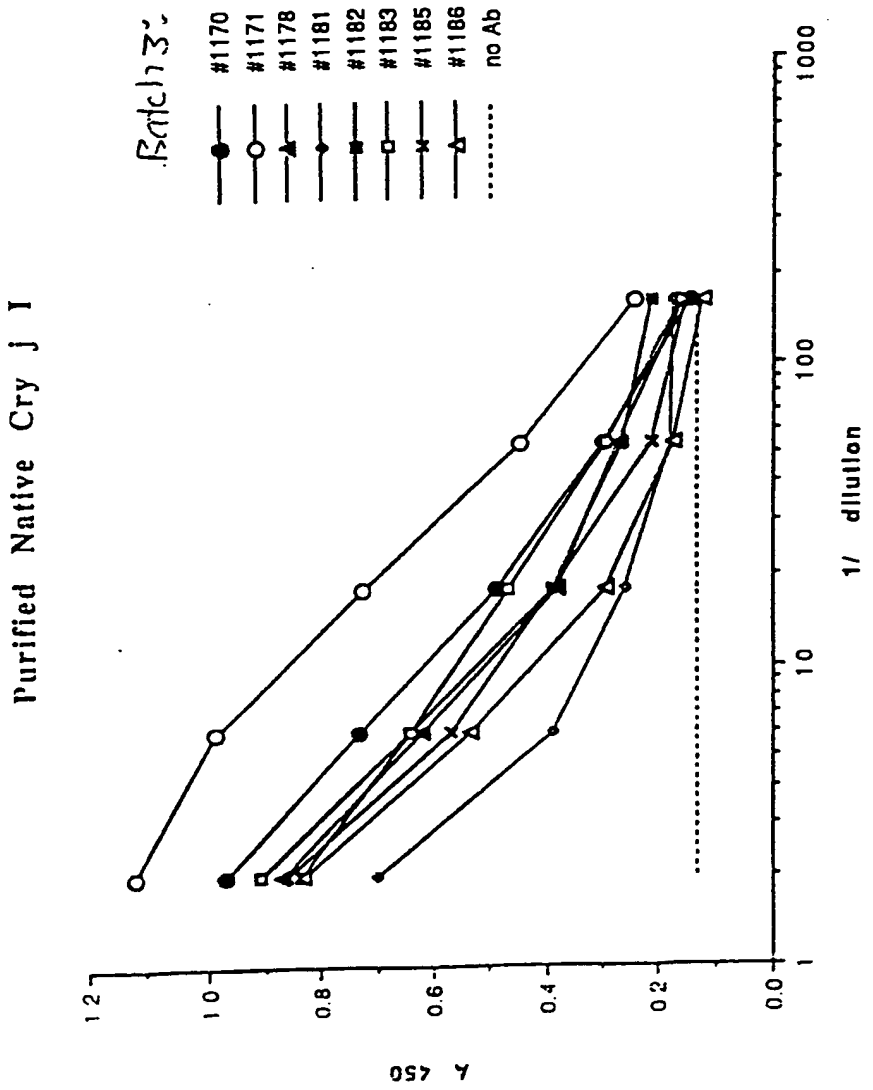
Purified Native Cry J I
Batch 3c
#1170
#1171
#1178
#1181
#1182
#1183
#1185
#1186
no Ab
A 450
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1
10
100
1000
1/ dilution

Fig. 13

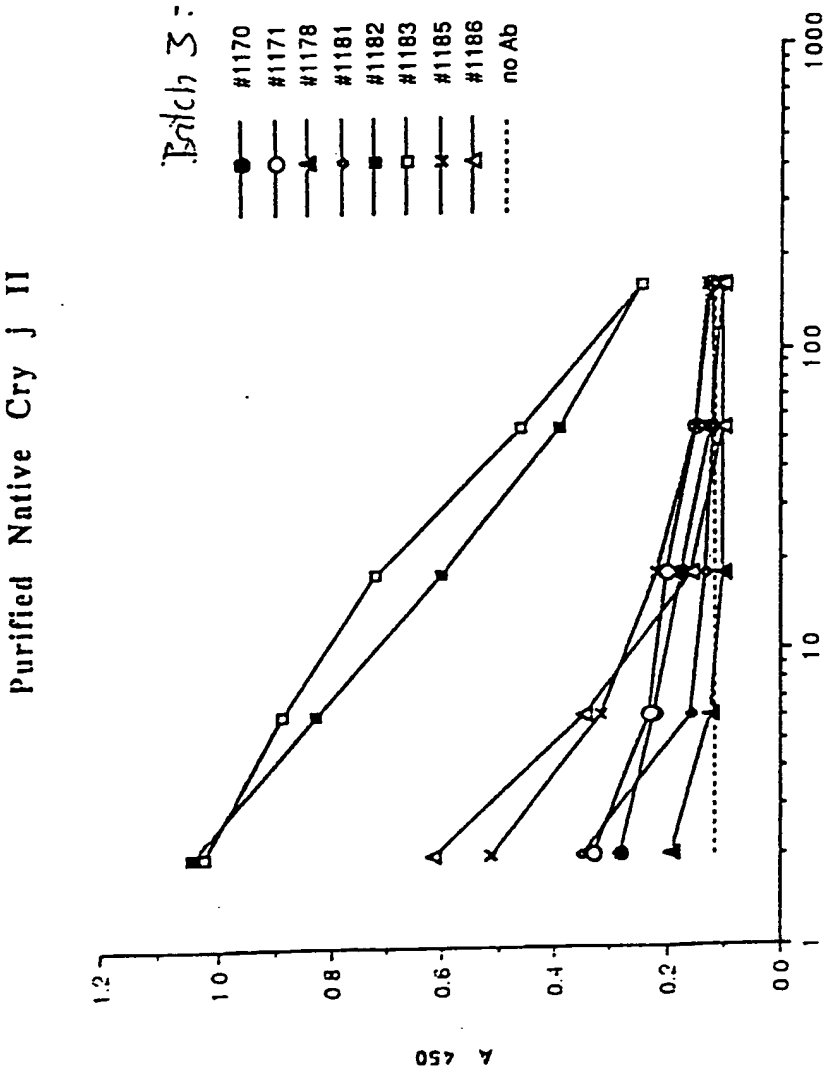
Purified Native Cry J II
Batch 3:
#1170
#1171
#1178
#1181
#1182
#1183
#1185
#1186
no Ab
A 450
1/ dilution
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1
10
100
1000

Fig. 14

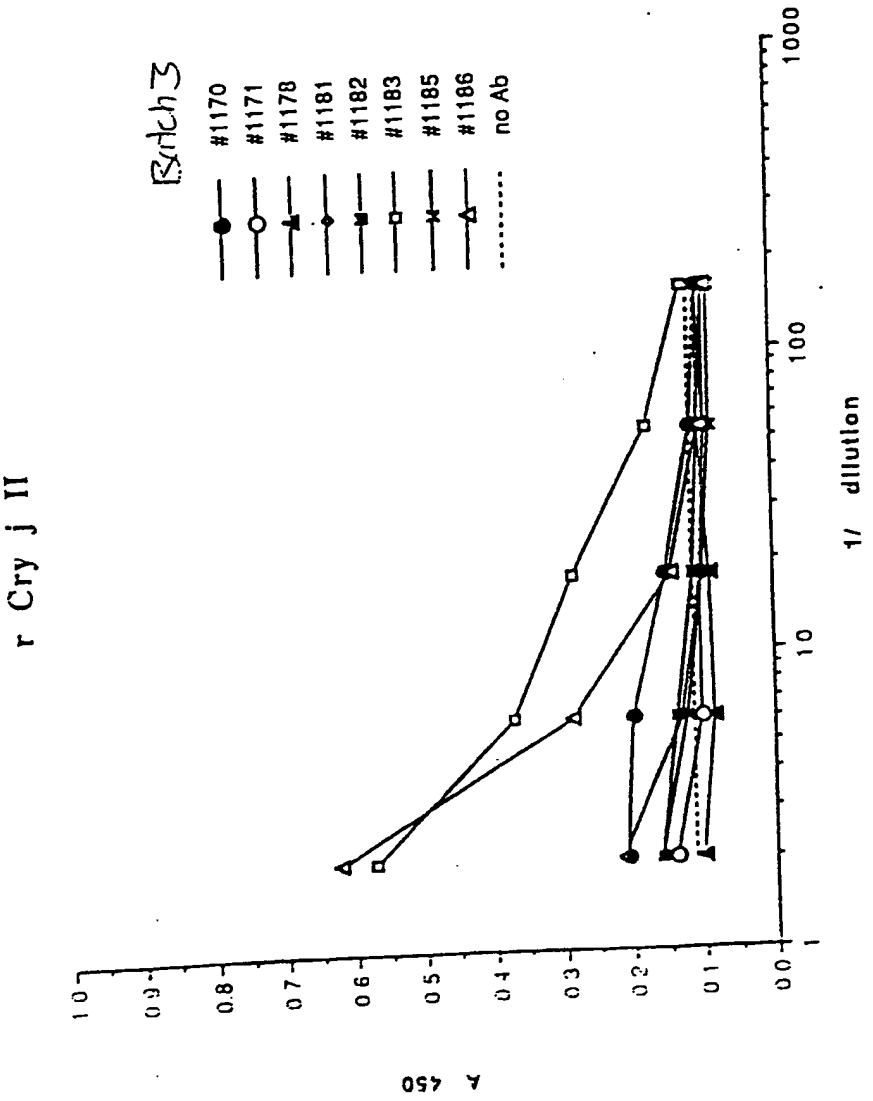
Rich 3
#1170
#1171
#1178
#1181
#1182
#1183
#1185
#1186
no Ab
r Cry j II
A 450
1/ dilution
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1
10
100
1000

Fig. 15

Fig. 16

Patient #	MAST	Purified Native *Cry j* I	Purified Native *Cry j* II	Recombinant *Cry j* II (r*Cry j* II)
1034	2	-	+	-
1142	2	+	-	-
1143	0	+	+	+
1144	1	+	+	-
1145	0	-	-	+
1146	3	+	-	-
1147	3	+	-	-
1148	3	+	+	+
1151	3	+	+	-
1153	1	+	-	-
1154	3	+	+	-
1158	2	+	+	-
1159	2	+	+	+
1165	1	+	-	+
1167	1/0	-	+	-
1170	1/0	+	-	-
1171	2	+	-	-
1178	1	+	-	-
1181	1/0	+	-	-
1182	1	+	+	-
1183	1	+	+	+
1185	1/0	+	+	-
1186	1/0	+	+	+
Positive	21	20	13	5

(g:1939.doc)